

Unterschrift des Betreuers



TECHNISCHE
UNIVERSITÄT
WIEN

Vienna University of Technology

DIPLOMARBEIT

Quantitative analysis of recombinant protein release using the Bacterial Ghost platform technology

ausgeführt am Institut für
Verfahrenstechnik, Umwelttechnik und technische Biowissenschaften
der Technischen Universität Wien

unter der Anleitung von
Univ.Prof.Dipl.-Ing. Dr.techn. Christoph Herwig und
Dipl.-Ing. Patrick Sagmeister

durch

Mag. Daniela Ehgartner Bakk.rer.nat.
Leystraße 134/21
1200 Wien

Datum

Unterschrift (Student)

Acknowledgements

I want to thank everybody who supported me in finishing this master thesis.

Especially I want to thank Prof. Christoph Herwig for the possibility to do the practical part of the master project in his working group. I really liked the atmosphere he has created in this division, where there is always someone who lends a hand.

Further I want to thank Patrick for the excellent supervision. There was always the possibility to be supported in no matter what situation – was it for questions concerning the fermentations, some help with the fermenter or topics concerning the writing of the thesis. I enjoyed the work together and I take a lot of knowledge and new ideas with me.

But also a big thank you to Andrea and Timo as additional supervisors, who were always there to give a good advice.

Also important were so many others in the laboratory, who answered questions, supported me to save my fermentation in critical situations or who helped me lifting the lid of my fermenter.

At the end I want to thank my family, my boyfriend and friends who supported me and who made a major contribution to the success of my study.

THANK YOU! DANKE! MERCI!

Daniela Ehgartner

Abstract

Bone morphogenetic protein (hBMP-2) is a bone growth factor applied on bone defects and for osteoporosis treatment. In recombinant production of hBMP-2 the downstream process is very time-consuming and costly. One common step in downstream processing is cell rupture, for instance homogenization. Bacterial ghosts are envelopes of Gram negative bacteria, which emerge by the formation of a lysis tunnel through the phage protein E. pBAD is an expression system providing triggering of protein production by the amount of L-arabinose in the medium and therefore avoiding high metabolic loads.

The aim of the study was to investigate the applicability of the bacterial ghost platform for the release of recombinant protein in a fermentation process. Furthermore, the applicability of a soft-sensor in process development was examined.

A DoE using specific D-glucose and L-arabinose uptake rates as critical process parameters was conducted. During induction phase a mixed feed system with D-glucose as the primary carbon source and L-arabinose as the secondary carbon source and inducer was applied. The specific substrate uptake rates were controlled by a first principle soft-sensor.

Lysis efficiency was optimal (99%) at high total specific substrate uptake rates and low concentrations of L-arabinose in the feed. The maximal bone morphogenetic protein titer ($5.6 \pm 1.5 \text{g/l}$) and the highest purity ($71 \pm 4\%$) were reached at a low total specific substrate uptake rate (0.1g/g/h). There was no difference between protein titers comparing homogenization and E-lysis as cell rupture methods. Purity was higher in homogenized samples, but E-lysed samples were purer than samples without cell rupture.

An optimal point for the fusion of bone morphogenetic protein production using the pBAD system and the bacterial ghost platform was found at 0.2g/g/h specific D-glucose and 0.05g/g/h specific L-arabinose uptake rate ($92.5 \pm 20\%$ lysis efficiency, $4.55 \pm 0.84 \text{g/l}$ protein titer). Although the physiological state of the cells at the end of production phase needs to be considered, the bacterial ghost platform provides a micro-biological alternative to classical mechanical cell rupture methods. The soft-sensor was successfully applied for the control of critical process parameters in the process development.

Content

Acknowledgements	2
Abstract	4
1. Introduction.....	10
1.1 Recombinant protein production in <i>E. coli</i>	11
1.1.1 The host strain.....	11
1.1.2 Promoter systems for <i>E. coli</i>	12
1.1.3 Molecular biology of <i>E. coli</i> : Cytoplasm, periplasm and extracellular space.....	12
1.1.4 High cell-density culture systems.....	13
1.1.5 Acetate formation	13
1.1.6 Metabolic burden.....	14
1.2 Target protein: hBMP-2	15
1.3 Expression system: pBAD	16
1.3.1 L-Arabinose transport and metabolism in <i>E. coli</i>	16
1.3.2 The pBAD expression system	18
1.3.3 Mixed feed systems.....	19
1.4 Ghost platform	21
1.4.1 Downstream process and BG application	24
1.5 PAT and QbD.....	25
1.6 Soft-sensors for process development.....	28
1.6.1 Approaches for soft-sensors	28
1.6.2 Hard-type sensor vs. soft-sensors.....	29
1.6.1 Process control using a soft-sensor.....	30
1.7 Multivariate data analysis	32
1.7.1 Multivariate data analysis in bioprocess technology.....	32

1.7.2	Rates and yield coefficients as physiological parameters.....	33
1.7.3	Interpreting yield coefficients and specific rates	34
1.8	Aims of this study	35
2.	Materials and Methods	36
2.1	Upstream process.....	36
2.1.1	The strain.....	36
2.1.2	Media.....	36
2.1.3	Bioreactor setup.....	37
2.1.4	Bioreactor instrumentation.....	38
2.1.5	Culture mode.....	39
2.2	Off-line analytics	41
2.2.1	Biomass dry cell weight.....	41
2.2.2	Extracellular protein determination.....	41
2.2.3	Homogenization and cell rupture	42
2.2.4	Purity determination of rhBMP-2 via SDS gel analysis.....	42
2.2.5	HBMP-2 titer investigation via RP-HPLC.....	42
2.2.6	Flow cytometry: Investigation of lysis efficiency	43
2.3	Multivariate data analysis	44
3.	Results and Discussion	46
3.1	Process Description and DoE Design	46
3.1.1	Soft-sensor assisted process control in combination with a mixed-feed system....	48
3.2	Data and information processing	52
3.2.1	Online data	52
3.2.2	Offline data.....	56

Extracellular protein	58
3.2.3 Processed data	59
3.2.4 Information matrix	63
3.2.5 Multivariate data analysis	63
3.1 E-lysis of <i>E. coli</i>	68
3.1.1 Flow cytometry: The method	68
3.1.2 Lysis efficiency	69
3.1.3 Lysis kinetics	73
3.1.4 Subpopulations during mixed feed phase 2	76
3.2 rhBMP-2 formation	78
3.3 Purity of rhBMP-2	82
3.4 Comparison of rhBMP-2 titer and purity in homogenized vs. lysed samples	85
4. Conclusions	89
4.1 Methodical conclusions	90
4.2 Physiological conclusions	91
5. Outlook	94
6. Appendix	97
6.1 Standard operation procedures	97
6.1.1 Protein precipitation	97
6.1.2 Extracellular protein measurement via BCA kit	99
6.2 Abbreviations in formulas	101
6.2.1 Symbols	101
6.2.2 Indices	102
6.2.3 Greek Symbols:	102
6.3 Data tables	103

6.4	Acquired raw and processed data	106
7.	References.....	137

1. Introduction

Today, over 200 protein pharmaceuticals like albumin, Factor VIII, human insulin and human growth hormone are produced worldwide [1]. In 2010 the sale of biological drugs reached worldwide sales of over US\$100 billion. The market is rising. Predictions expect that more than 50% of new drug approvals will be biological drugs in 2015. Until 2025 a raise to more than 70% is predicted [2].

Most of these proteins are synthesized in recombinant expression systems like bacteria, yeasts, plants, mammals, molds or insects [1]. Recently also transgenic plants and animals were used like the production of human α_1 -proteinase inhibitor in transgenic tomato plants [3] or Antithrombin III in the milk of transgenic mammals [4].

In this study the therapeutic protein recombinant human bone morphogenetic protein 2 (rh BMP-2) is produced using *E. coli* as expression system. Typically, recombinant protein products need to be liberated by cell rupture methods like high pressure homogenization in *E. coli* bioprocesses [5]. To facilitate downstream processing of the product, the ghost technology platform provides a possibility to release the recombinant protein in the surrounding medium. The ability to lyse of the ghost platform is dependent on physiological state of the cells [6, 7]. The pBAD mixed feed expression system enables the control of recombinant protein production by making use of a mixed feed-system, where the specific uptake rate of arabinose in the feed is proportional to induction state [8].

Within this study, the impact of specific L-arabinose uptake and D-glucose uptake on process productivity and recombinant protein release efficiency is investigated. Multivariate methods are applied on the data to explore not only the interactions between CPPs and CQAs, but to further investigate in physiological variability of the organism which links CPPs and CQAs. These physiological parameters are represented by yield coefficients and rates. Thereby process understanding is enhanced.

1.1 Recombinant protein production in *E. coli*

1.1.1 The host strain

Escherichia coli, a gram negative bacterium, is the primary microbial host [9-11] by producing 39% of all recombinant proteins on the market [1]. In concern to physiology, molecular genetics and expression systems it is the best known organism. Additionally rapid growth and high protein yields are key advantages of this bacterium [1, 11, 12]. Complex proteins often depict problems in protein folding and post-translational modifications [1, 12]. Therefore, when modifications such as the post-translational glycosylation are required, eukaryotic systems like fungi and mammalian cells are more suited [1]. But if recombinant proteins are required without complex post-translational modifications usually *E. coli* is selected as host [13].

To generate host organisms for recombinant protein production, the target DNA first has to be cloned and then is transfected to be amplified into the chosen expression system [1]. A recombinant expression system needs to include transcriptional promoters, origin of replication, translation initiation markers, transcriptional and translational terminators as well as an antibiotic resistance marker. Common expression systems in *E. coli* are the pET and the pBAD expression systems [14].

Furthermore, codons chosen in the heterologous genes are better chosen dependent on the codon usage of *E. coli*, because this may be a limiting factor in recombinant protein formation. There exist 61 amino acid codons for the production of mRNA molecules. Problems in translation occur, when codons not typical for *E. coli* appear. Hence heterologous genes containing untypical codons for *E. coli*, may be inefficiently expressed [15].

In this study *E. coli* C41 is used. It is a mutant of the strain *E. coli* BL21 (DE3), for which recombinant protein production is often toxic. *E. coli* C41 was found by overexpressing recombinant proteins in *E. coli* BL21 (DE3). The mutants not died by toxic effects, were collected and examined. The difference of *E. coli* C41 (the mutant) to *E. coli* BL21 (DE3) is a lower transcription of a transport protein from mitochondria, which seems to be distinct for the production of high recombinant protein amounts [16]. Furthermore, *E. coli* C41 is able to metabolize L-arabinose [17], as necessary for the use in a mixed feed system.

1.1.2 Promoter systems for *E. coli*

Typical promoter systems used in *E. coli* are *lac*, *tac* and *trc* [1]. The systems need to be simply and inexpensively inductive, whereby the inducers should be independent from common media ingredients [1, 15]. Furthermore, a tight regulation is seen as advantage as a low-basal level of expression is desirable to minimize toxic effects and metabolic burdens [15]. Dependent on the actual requirements the promoter is chosen. For high expression which should be easily inducible, *tac* or *trc* are the most recommended ones [18]. These strong promoters are commonly used for shorter recombinant protein production period, as the product formation rate rapidly decreases [19].

When the target protein is potentially insoluble or toxic, or when other complications due to a too high protein expression could occur, *araBAD* can be chosen, as the level of induction is dependent of arabinose concentration in the feed [18].

1.1.3 Molecular biology of *E. coli*: Cytoplasm, periplasm and extracellular space

E. coli has an inner and an outer membrane which separate cytoplasm, periplasm and extracellular space. In the cytoplasm the highest protein yield can be achieved [1, 11], but inclusion body formation might happen due to an imbalance between in vivo protein aggregation and protein folding [14, 20]. Hydrophobic patches of non- or only partially folded protein form aggregates [20]. At macromolecule concentrations in the cytoplasm up to 300-400 mg/ml, the protein folding is a challenge especially for large overexpressed heterologous proteins which require chaperons as folding helpers. Inclusion bodies in the cytoplasm can have a volume up to $0.6 \mu\text{m}^3$ [13] and a diameter of more than $1 \mu\text{m}$ [21].

Aggregation can be decreased by controlling parameters like expression rate, temperature, host metabolism as well as target protein engineering such as co-expression of plasmid-encoded chaperones and solubility tag-technology [14]. To receive the final product out of inclusion bodies, a protein denaturation and refolding process is needed [1, 11].

But inclusion body formation also has advantages, as inclusion bodies can accumulate in the cytoplasm in higher quantities as soluble proteins do and proteins thereby can be isolated in a highly concentrated and purified state [11, 13]. Furthermore, inclusion bodies are resistant to *E.*

coli proteases and due to this protein inactivity toxic products for the bacteria can be expressed [11, 13].

Periplasmic production (protein storage in the periplasm) normally prevents inclusion body formation [1, 11, 22]. Additionally protein folding is improved as the oxidative environment enables disulfide bond formation [11, 12, 22], the proteins are prevented from protease degradation and due to a lower native protein content in the periplasm, purification of the product is facilitated [1, 11].

Intended excretion of recombinant proteins into the medium is not very common and offers some difficulties like inefficient secretion [22]. It simplifies purification as cell rupture is not required to release the product. Furthermore, proteins are more soluble in the medium than in the cytoplasm and they are prevented from protease degradation, which leads to higher stability [11, 22]. However, the product is diluted in the medium and the yield is low compared to production into the cytoplasm [11].

1.1.4 High cell-density culture systems

When high cell-density culture systems like batch or fed-batch are run, high cell concentrations can be provided (100 g dry cell weight/liter in fed batch) [15, 23].

Fed-batch is most common in recombinant protein production as by substrate limitation engineering shortcomings in cooling and oxygen transfer can be avoided. Furthermore the substrate limitation renders a sort of metabolic control to avoid overflow metabolism and thereby by-product formation [23]. An exponential feed provides a controlled constant growth rate of cells. Therefore a cost-effective production of proteins can be achieved [23, 24].

1.1.5 Acetate formation

Acetate formation is a by-product in *E. coli* which is produced due to a limited oxidative capacity for carbon source metabolism [1, 9, 25]. Acetate formation is inversely correlated to dissolved oxygen levels in the culture broth [26]. It is directly linked to growth rate respectively substrate uptake rate. Up from a threshold level of growth rate, the capacity for aerobically degradation is

limited by the maximal oxygen consumption rate of the bacteria, and acetate is formed as by-product [25]. This phenomenon is also called the bacterial Crabtree effect [27]. Especially in recombinant-protein producing cells acetate retards cell growth by decreasing rate of DNA, RNA, lipid and protein synthesis. Additionally, the carbon used for acetate formation would be used otherwise to build biomass. Furthermore, acetate affects proteins and genes involved in stress response, transcription-translation machinery and regulation [25, 28]. Acetate production can be avoided by controlling oxygen levels [1], developing appropriate feeding strategies to control substrate uptake [1, 25] or genetic modification [9, 25].

1.1.6 Metabolic burden

The so-called metabolic burden is an alteration in the carbon metabolism of the host cells, induced by recombinant protein production. Alterations occur due to the shift from the utilization of the carbon source for biomass and by-product production to energy formation [29]. Energy is needed for translation machinery synthesis, translation, and amino acid synthesis for recombinant protein production [10, 29]. The mechanisms to shift carbon source utilization from biomass production to energy formation is by increasing conversion of NADPH to NADH for ATP generation and by the redirection of cell fluxes towards acetate and acetyl-CoA to form additional NADH and ATP [29].

Carbon or amino acid depletion leads to inhibited RNA and protein synthesis which is mediated by the alarmone guanosine tetraphosphate (ppGpp) [23, 30]. This so called stringent response works by the binding of ppGpp to the RNA polymerase core enzyme whereby gene expression is affected. A down-regulation of genes involved in growth and cell proliferation occurs on the one hand, and an up-regulation of genes concerning maintenance and stress defense on the other hand [23].

The metabolic burden can be overcome by genetic engineering of the host organism or by appropriate cultivation conditions such as amino acid supplementation through the medium, low glucose concentration in the medium as well as the control of the specific growth rate [10]. The stress response due to recombinant protein production is linked to the rates of transcription, but also can be raised by specific properties of the protein like misfolding [23].

It is not only recombinant protein production which imposes a metabolic burden. The plasmid maintenance alone already results in a decreased growth rate and cell density, when comparing to organisms not bearing the plasmid [31, 32]. This is due to perturbed gene regulation namely by down-regulation of energy metabolism genes, up-regulated heat shock proteins and a difference in transport gene expression [31].

Furthermore, aggregation and protein misfolding of intracellular recombinant proteins influences the cell response. It affects host proteins concerning oxidative stress and lipid metabolism, rearrangement of membrane lipid composition as well as it decreases the permeability and the fluidity of the membrane. Additionally aggregates induce the expression of proteins like proteases to remove the aggregated proteins [33]. But also chaperons and other folding catalysts are higher expressed which can increase formation and folding of the produced recombinant protein [9, 33].

1.2 Target protein: hBMP-2

Bone morphogenetic proteins (BMPs) were first discovered in an osteoinductive extract of bone matrix [34]. These proteins are growth factors of the growth factor- β superfamily which are able to de novo form bone tissue [34, 35]. In 2002 the FDA approved BMPs to be applied in humans [36].

BMP-2, with the size of 14 kDa [37], is a homodimeric protein whereas the two monomers hold cysteine knots where the polypeptide backbone is connected to a ring by two disulfide bridges. A third disulfide bond is interlinked to the ring. By this disulfide pattern stabilized monomers are arranged head-to-tail to build the dimer which is held by intermolecular disulfide bonds and hydrophobic interactions [36-38].

Members of the BMP family which actually are in clinical use are recombinant human (rh) BMP-2 and rhBMP-4 [34]. BMP-2 can be used to heal bone defects as well as an application as treatment of osteoporosis is probable. For spinal fusion surgery clinical studies have shown that BMP-2 used as bone graft substitutes has reasonable effects. Furthermore, dental medicine can profit of BMP-2 by using it to induce new dentine formation as well as a substitute for root canal surgery [39]. The commercial product in which BMP-2 is actually available is called InductOs in

use as an alternative to bone autograft [36]. The actual prize for 1mg of lyophilized rhBMP-2 powder produced by *E. coli* is \$4680 [40].

RhBMP-2 produced by mammalian cell expression has the advantage of receiving processed, dimerized and glycosylated molecules [34, 41]. However, low yields and incomplete monomer processing are negative sides in this kind of manufacturing [36, 41]. Eukaryotic expression systems used beside mammalian cells are transgenic tobacco plants. The production process in plants is time intensive. Additionally a four-step purification protocol is necessary [36].

Furthermore, *E. coli* is used as host for the production of rhBMP-2 [36, 41, 42]. As for BMP-2 glycosylation is not crucial, prokaryotic hosts offer a possibility of producing BMP-2 more effectively [36]. Solubilization and refolding are critical steps in this process [36, 41, 42] as they are crucial for recovery of biological activity of the protein. The highest influence on renaturation yield have temperature, presence of aggregation suppressors, pH and protein concentration [37].

1.3 Expression system: pBAD

1.3.1 L-Arabinose transport and metabolism in *E. coli*

In *E. coli* L-arabinose is metabolized by the proteins of *araBAD* to D-xylose-5-phosphate to enter the pentose phosphate shunt [43]. When a high level of glucose as the preferred carbon source is present in the medium, the so-called catabolite repression occurs by inhibiting gene expression and protein activity caused by low levels of cyclic adenosine monophosphate (cAMP) to prevent the uptake of other carbon sources like L-arabinose [23].

The gene-enzyme complex for L-arabinose in *E. coli* consists of the structural genes *araA*, *araB* and *araD*, the controlling site *araI* as initiator and the positively and negatively regulating gene *araC*. Additionally there exists the gene *araE* which has been shown to be responsible for the active (low-affinity) transport of L-arabinose [44-46]. It is bound to the inner membrane and transports L-arabinose using the electrochemical potential [47]. The metabolism of L-arabinose requires apart of *araBAD* and *araC* the cAMP receptor protein CAP [48] because this protein either stimulates binding of RNA polymerase or the translation to the open complex as well as it facilitates to open the DNA loop which is built when arabinose catabolism is repressed [47].

Furthermore, the *araFGH* operon includes the genes for the L-arabinose-binding protein (AraF) and two membrane-associated proteins AraG and AraH which are concerned with high-affinity transport of L-arabinose [46, 49]. These are working like an ABC transporter system [47].

The AraC protein regulates the L-arabinose operon positively and negatively dependent on the concentration of L-arabinose. It consists of a DNA-binding domain and an arabinose binding pocket, which are linked. The protein builds dimers (see Figure 1a). In the absence of L-arabinose (see Figure 1b), one part of the AraC dimer occupies the *araI*₁ site, while the other one is occupying the 210 base pairs distant site called *araO*₂ by DNA looping. Hence the binding site at *p*_C is no more accessible for the RNA polymerase [43, 46, 50-52]. Furthermore the RNA polymerase has no access to the *p*_{BAD} promoter and also the cAMP receptor protein might be hindered to bind alongside AraC [43].

When L-arabinose is added, the two monomers of AraC occupy the adjacent half-sites *araI*₂ and *araI*₁ whereby *p*_{BAD} is induced as well as *p*_C is now accessible for RNA polymerase (see Figure 1c) [43, 46]. Additionally to AraC the cAMP-catabolite activator protein influences the regulation of these promoters. The response of the operons to arabinose takes 15 to 30 sec [46].

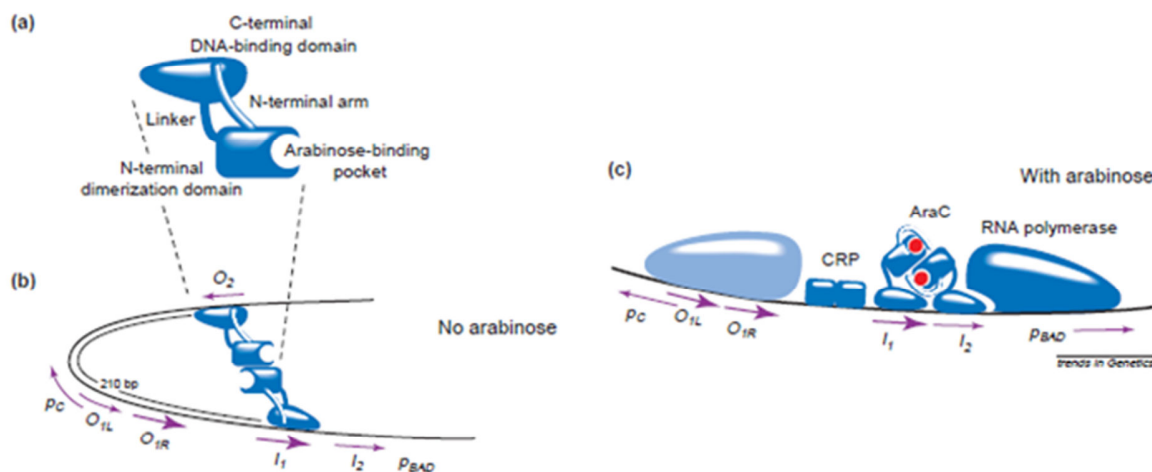


Figure 1: Regulation of the arabinose operon via L-arabinose (adapted from [43])

1.3.2 The pBAD expression system

The plasmid pBAD, which is used in recombinant protein production, is based on the promoter p_{BAD} and its regulator *araC*. This expression system enables high expression of target protein at the presence of inducer. Furthermore the height of expression can be modulated by the range of inducer concentration (see Figure 2 on the left). Thereby problems associated with high recombinant protein production can be avoided. A further advantage of pBAD is that before induction the system is well repressed [53]. Additionally an actual expression can be repressed by catabolite repression via high amounts of glucose or by competitive binding of fucose, which functions as anti-inducer. Hence protein production can rapidly be turned off by varying the sugars in the medium [54]. As well as a fast stop of protein production, induction rate also happens within some minutes [53]. When the recombinant protein is expressed via the pBAD system, a mixed-feed approach consisting of arabinose as secondary carbon source to regulate recombinant protein production can be advantageous (see below) [55-57].

Sagmeister et al. [57] applied a mixed-feed system for *E. coli* containing pBAD to investigate the maximal specific uptake rates of the inducer L-arabinose as function of the specific uptake rate of glucose. A combination of rate-based first principle soft-sensor and online infrared spectroscopy was used as control strategy of substrate feed. The soft-sensor calculated the biomass concentration depending on flow rates whereas infrared spectroscopy measured residual substrate concentration in the fermenter. The maximal specific L-arabinose uptake rates as function of specific glucose uptake rates are shown in Figure 2. The cell is able to metabolize more than a total specific substrate uptake rate (sum of specific D-glucose and L-arabinose uptake rate) of 1 g/g/h, without by-product formation [57].

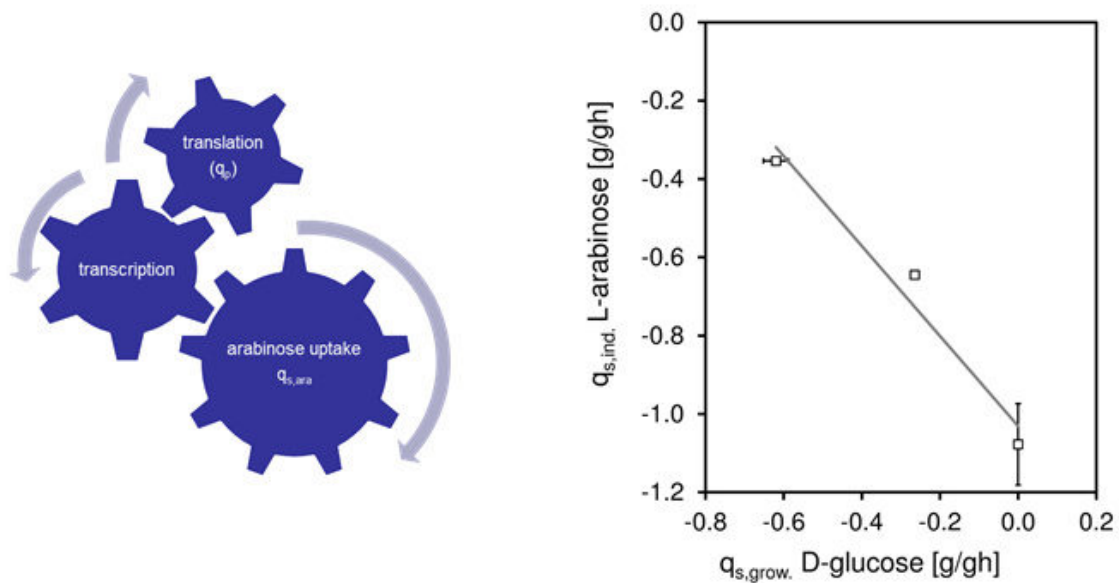


Figure 2: Scheme of expression regulation by inducer addition (left) and the dependency of specific L-arabinose uptake rate on specific glucose uptake rate (adapted from Sagmeister et al. [57])

Even though L-arabinose concentration modulates protein expression over the whole cell population, p_{BAD} is expressed in an all-or-non fashioned manner for each single cell. Thus the amount of induced cells varies with the concentration of L-arabinose in the medium [58]. Expression in an all- or non-fashion occurs due to the fact that the expression of L-arabinose transport genes as well as L-arabinose metabolic genes is all dependent on the same inducer. Hence, only in cells which are able to transport enough inducer inside, more transporters are synthesized to even raise the arabinose level in the cell. However, cells with transporter levels beneath a certain threshold are not able to transport enough L-arabinose to be induced [59-61]. To face this problem the low-affinity high-capacity transporter (AraE) which was inducible independently of arabinose was expressed. Thus the induction of p_{BAD} in each cell showed a linear response to the arabinose concentration in the medium [59, 60].

1.3.3 Mixed feed systems

Mixed-feed systems consist of two different carbon sources which are fed in a determined ratio. These systems are applied in the mean that the first carbon source is linked to cell growth and

cell energy, whereas the second carbon source is used to regulate recombinant protein production [57].

Similar to the p_{BAD} system, the alcohol oxidase promoter in *P. pastoris* is induced by a probable carbon source of the organism. Activation of protein production thereby is done by feeding methanol, which in parallel is used as a secondary carbon source. The primary carbon source glycerol which is applied for continuous energy and growth supply – like glucose in p_{BAD} – works as repression/depression of the promoter. Mixed feed systems in a fed-batch culture have shown to increase productivity of heterologous proteins. The concentration of the inducer is chosen to be high enough to fully induce recombinant protein production, but to be low enough to prevent growth inhibition and by-product formation [24]. A study of Jungo et al. (2007a) [55] showed that mixed substrate feed (glycerol and methanol) in induction phase in *P. pastoris* resulted in an increased volumetric recombinant protein production and a higher biomass yield. Furthermore decreased heat production and oxygen uptake rate brought technical advantages [55]. To avoid the repression of alcohol oxidase by glycerol, a mixed-feed approach for *P. pastoris* can be done using sorbitol instead [62]. When still glycerol is the primary carbon source, carbon-limited growth conditions enable a simultaneous decomposition of glycerol and methanol instead of sequential consumption [55]. Hence the control of the glycerol feeding rate is an important parameter. When glycerol feeding rate respective to biomass should be held constant, the actual biomass concentration needs to be known. Therefore, instead of the normally used specific growth rates, the specific uptake rate was applied to calculate current feed rates. This parameter suited better due to a deviant biomass yield during recombinant protein production compared to growth phase [63].

Recently mixed feed systems were also applied in *E. coli* to reduce the metabolic burden caused by too high recombinant protein production [8]. For recombinant protein expression the p_{BAD} system was used with a mixed feed consisting of D-glucose as primary and L-arabinose as secondary carbon source. Additionally L-arabinose was used as an inducer. To render metabolism of L-arabinose possible, an *E. coli* C41 strain was used. Tuning of the transcription rate on cellular level was demonstrated using flow cytometric analysis with the model product green fluorescent protein (GFP). Beside of the prolonged expression time going in line with

reduced metabolic burden, the ability of triggering recombinant protein production rate by the amount of inducer in the feed was shown [8].

1.4 Ghost platform

Bacterial ghosts (BGs) are envelopes of Gram-negative bacteria emerged by controlled expression of the cloned lysis gene *E* [6]. These envelopes possess all surface structures of the living cell like adhesins, outer membrane proteins and pili presented in a native configuration [6, 64].

Gene *E*, derived from phage ϕ X174, which is coding for a 91-aa polypeptide, has a bacteriophage origin. Protein *E* is a membrane protein which is able to build a transmembrane tunnel structure by fusing inner and outer membranes (see Figure 3; [65, 66]). Border values of the tunnel are between 40 and 200 nm [6]. Tunnel formation is described by Schön et al. [67] to happen in three phases. First the protein *E* is integrated into the inner membrane. Furthermore, the C-terminus which has been facing the cytoplasm is transferred to the periplasmic space by conformational change of the protein. Additionally an oligomerization takes place and the division initiation complex is targeted. The last phase includes the exposition of the C-terminus of protein *E* to the cell surface, which induces the fusion of inner and outer membrane at membrane adhesion sites to form a tunnel.

Due to the osmotic pressure difference between the cytoplasm and the bacterial environment, the cell content is discharged. In contrast, the content of the periplasm rests in the BG cell as by fusion of inner and outer membrane the periplasmic space is sealed [6].

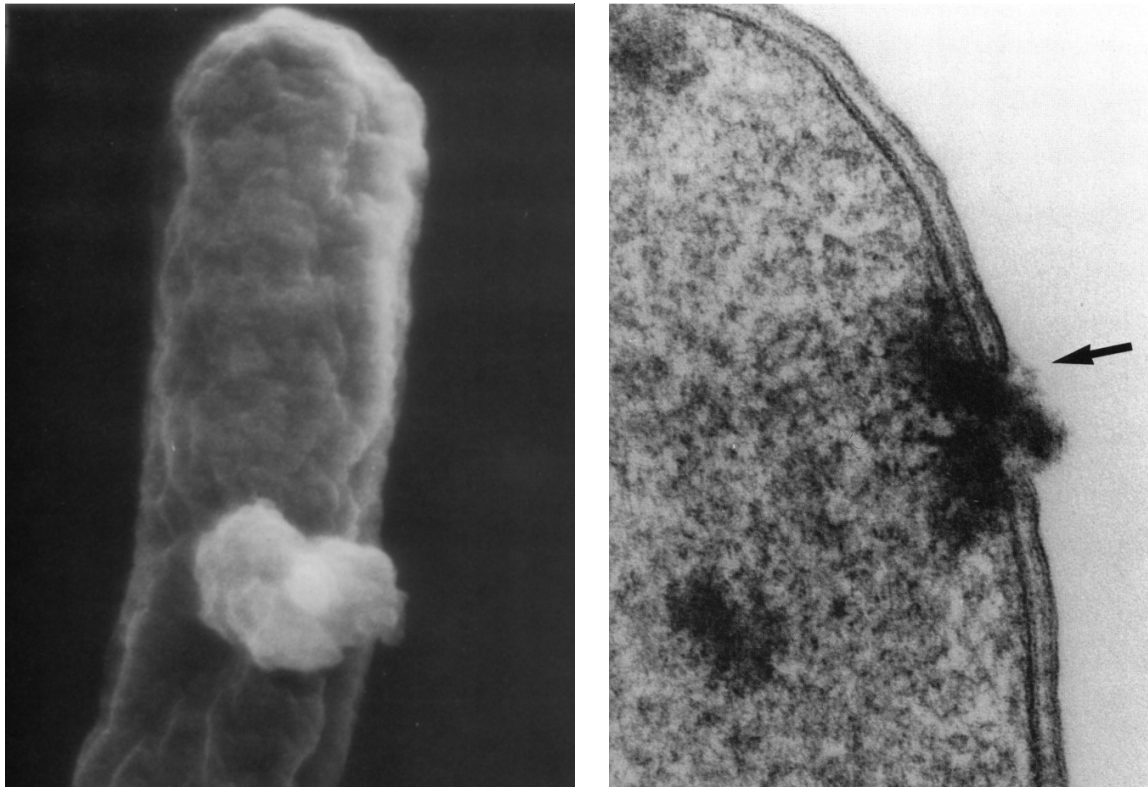


Figure 3: E.coli shortly after E-lysis (adapted from Witte et al. [66])

To produce bacterial ghosts, the cells need to be transformed with the plasmid using the gene for protein E. Expression systems used are the *lac* and temperature-sensitive Lambda promoter/operator systems [68]. Actually especially an optimized Lambda system is applied where E-mediated lysis is induced at temperatures of 42 or 44°C [7]. Apart from the plasmid for E-lysis, BGs used for the release of recombinant protein have a second plasmid inserted, carrying the gene for heterologous protein production. Hence protein production and E-mediated lysis are induced separately.

The lysis tunnel is located either at the poles or at the center of bacteria, which both are possible division sites. Therefore the assumption was made, that the ability to lyse is dependent on the physiological state of the cell and its ability to divide [6, 7]. Recent research showed that E-lysis is dependent on the physiological state of bacteria. A specific D-glucose uptake rate higher than 0.4g/g/h is distinct for the ability to lyse. When the substrate is limiting, no E-lysis occur, even when protein E is present. The same is true for the impact of temperature on the E-lysis capable state of the cells, as only above a certain temperature threshold (24°C) an

appropriate D-glucose uptake rate and hence a certain metabolic activity is provided [69]. Furthermore, a non-physiological pH showed to disturb E-lysis [7].

Therefore, fermentation conditions should be controlled to not limit E-lysis kinetics. Due to the metabolic burden caused by recombinant protein production the metabolism changes and viability of bacteria can be decreased [56]. Thus recombinant protein production before lysis needs to be regulated, to not extensively stress the cell and to retain the culture in a lysis competent state.

There are manifold applications of BGs in progress which are further explained:

- **BGs** can be used **solo** to immunize against pathogenic Gram-negative bacteria. This has been studied in various animal models by using BGs as vaccines for *A. pleuropneumoniae* to prevent aerogenic infection. Furthermore, clinical studies were conducted for *V. cholera* on rabbits, whereas cross-protections were detected. Hence the animals were not only immune against the classical strain, but as well against new ones [6].
- A further application for this platform is **BGs as adjuvants**. To stimulate and potentiate the immune response to a target antigen, adjuvants like Cholera toxin, lectins, lipid A, muramyl dipeptide and others can be used [6, 70]. One possibility to prepare this application of ghosts is to mix a suspension of bacterial ghosts with a foreign adjuvant [6].
- By incorporation of foreign antigens into the envelope complex, BGs can function as **carriers of foreign protein antigens**. These antigens are fused either with the inner or the outer membrane, and are thereby presented. Furthermore, there exists a method to incorporate foreign proteins by transporting them to the periplasmic space, which is sealed during lysis, whereas proteins accumulated there, are retained in the bacterial ghost [7, 71]. A big advantage of using bacterial ghosts as carrier of antigens is that no inactivation methods are required which would denature immunogenic determinants [68].
- In white biotechnology BGs have their functionality in working as **micro-bioreactors for enzymatic reactions**. Bacterial lysis does not mean that all enzymatic activity is eliminated. Membrane-bound β -galactosidase, chloramphenicol acetyl transferase and

further enzymes have been described to still be active in BGs. This application of BGs can be used for the production of fine chemicals like asymmetric synthesis of 3,5-dicarboxyhydroxylate in biphasic ionic liquid systems [7].

- BGs produced as **carriers of DNA vaccines** are regarded as safer as conventional viral and bacterial vaccine delivery systems, because they have neither genotoxic nor cytotoxic impact on human cells after mutual incubation [7, 72]. Ebenson et al. [73] evaluated *Mannheimia haemolytica* as ghosts as delivery system for DNA vaccines. Results of this study show high transfection efficiencies which have not been able to be achieved by conventional naked DNA vaccines [73].
- **BGs as carrier for biologically active substances:** The carrier capacity of the cytoplasmic space is 250 femtoliter per BG [7]. Paukner, Kohl and Lubitz [74] showed that BGs are an appropriate drug delivery system for the drug doxorubicin, which is associated with the bacterial ghosts. Ghosts release the drug slowly at the target cells and are thereby a successful application in cancer treatment [74]. Apart from carrying substances for clinical use, BGs are effectively employed as carrier system for pesticides [75].

1.4.1 Downstream process and BG application

A new application of E-mediated lysis is BG production for the release of recombinant protein. Thereby the downstream processing of bioprocesses can be facilitated, as intracellular content is released into the surrounding medium. Alternative cell rupture methods to release the protein from the cell are detergent treatment, cell rupture by lysozyme, sonication, freeze-thaw [18] but as well high-pressure homogenization, bead milling and thermolysis, which means to heat the fermentation broth to 60°C and higher [5].

A conventional downstream process like it could be applied for rhBMP-2 produced in *E. coli* is shown in Figure 4. In the fermentation broth the whole (and closed) cells are suspended in the medium. Through centrifugation, the cells are separated from the medium. A cell ruptures method like homogenization is necessary to open the cells, to be able to separate the cytoplasm from inclusion bodies and cell debris. In the case of rhBMP-2 a solubilization is required to bring the hydrophobic protein in solution.

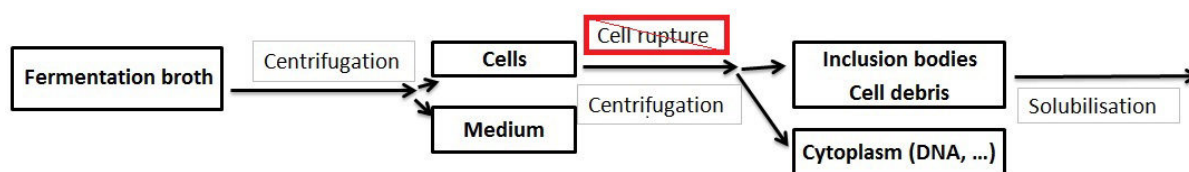


Figure 4: Conventional down-stream process for inclusion body processes.

In contrast to the above described down-stream process, the execution after a recombinant protein process applying the bacterial ghost technology is different. In this case the fermentation broth consists of opened cells hence a composition of medium and cell content. Through centrifugation inclusion bodies and cell debris can directly be separated from the cytoplasm. The inclusion bodies and the cell debris, which consists of the bacterial envelope where the periplasmatic space is sealed, are solubilized to receive rhBMP-2 in solution.

By avoiding the cell rupture step, time and money is saved in the down-stream process.

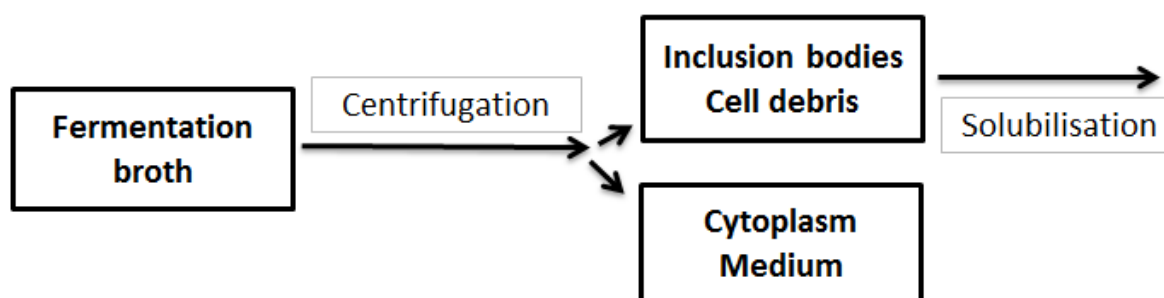


Figure 5: Section from a down-stream process where the bacterial ghosts technology had been applied at the end of fermentation

1.5 PAT and QbD

Since 1962 the good manufacturing practices (GMP) is a legal standard to assure that all drugs are adequately prepared and that the manufacturing procedures are controlled [76]. To improve the pharmaceutical manufacturing process, the American food and drug administration (FDA) launched a new initiative in 2002 to guarantee the access of American public to quality health care services. Furthermore, the ICH (International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use), which is supported by the FDA, introduced process analytical technology (PAT) into the pharmaceutical production and quality control [35]. The ICH defined PAT as “a system for designing, analyzing, and

controlling manufacturing through timely measurements (i.e., during processing) of critical quality and performance attributes of raw and in-process materials and processes with the goal of ensuring final product quality” [35]. To introduce PAT in Europe, the European Medicines Agency (EMA) established an EMA PAT team [76].

The aim of PAT is to increase process understanding and controlling. The approach is that product quality should be designed by a controlled and understood manufacturing process. In detail an understanding of intended therapeutic objectives concerning e.g. the patient population and the route of administration, the chemical, physical and biopharmaceutical characteristics of the drug, the design of the product as well as the design of the manufacturing process should be given [35].

Quality by Design (QbD), which is closely linked to PAT, includes the overall approach to development “that begins with predefined objectives and emphasizes product and process understanding based on sound science and quality risk management” [36]. It bases on entire knowledge of product and process which include the understanding of interactions between CQAs and the process, the relationship between critical quality attributes (CQAs) and the clinical properties of the product as well as a knowledge of the variability in raw materials [77].

Understanding of a process means that critical sources of variability – critical process parameters (CPP) – are monitored and explained. Hence this variability needs to be controlled by the process and a design space should be established for the conditions of the manufacturing process, which ensures the accurate prediction of CQA [36].

The design space links input variables and process parameters with CQAs [38]. CQAs are “physical, chemical, biological or microbiological property or characteristic that should be within an appropriate limit, range, or distribution to ensure the desired product quality”. An iterative process of experimentation and quality risk management can be used to prepare a list of CQAs [36].

Risk assessment is a structured approach to consider critical areas due to the impact of process parameters to determine further steps. As a rule, risk assessment is performed at the beginning of pharmaceutical development. But it can also be repeated when a greater knowledge of the process becomes available. A list of potential parameters is created whereof the significant ones

are identified and further studied by using for example a combination of mathematical models and design of experiment (DoE) [37].

A DoE can be used to investigate the multivariate interactions of CPPs and how they are linked to product quality and productivity. The thereby created design space includes the influence of multivariate functional interaction and combinations of process parameters and input variables like material attributes on product quality [36]. Variation of parameters within the design space does not lead to changes in quality, which enables to guarantee product quality when parameter changes due to scale up or manufacturing faults occur [36]. A design space is defined for the product, but later on the process as well. When the acceptable variability in CQAs from the product design space is determined, an acceptable variability of process parameters is defined using process characterization studies [77]. For process validation the CQAs and CPPs have to be included in a control strategy which is derived from product and process understanding, to assure product quality and process performance [78].

Figure 6 shows the connection of process and product knowledge with design space and normal operating ranges. The design space lies within the knowledge space, which contains information about non-critical and critical attributes well as process parameters. This knowledge is obtained during product development and all regions included have been investigated. Unexplored space is excluded. Areas of the knowledge space which are known to have a good product quality are placed inside the design space, whereas regions with unacceptable product are outside. The normal operating ranges are placed in the design space [78].

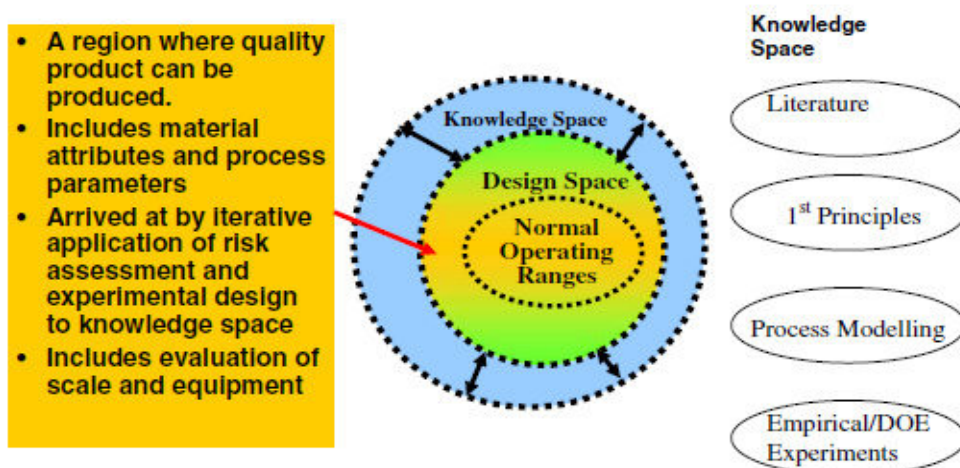


Figure 6: Illustration of design space (adapted from Lepore and Spavins [79])

1.6 Soft-sensors for process development

1.6.1 Approaches for soft-sensors

Soft-sensors are algorithms which calculate estimated values of process variables which are not measurable in real time. The estimation is based on knowledge about model structure and fermentation parameters. It is calculated using online available data like gas rates, reactor weight and substrate feed [80, 81]. The principle of a soft-sensor is the association between a sensor (hardware) and an estimator (software) (see scheme in Figure 7) [82].

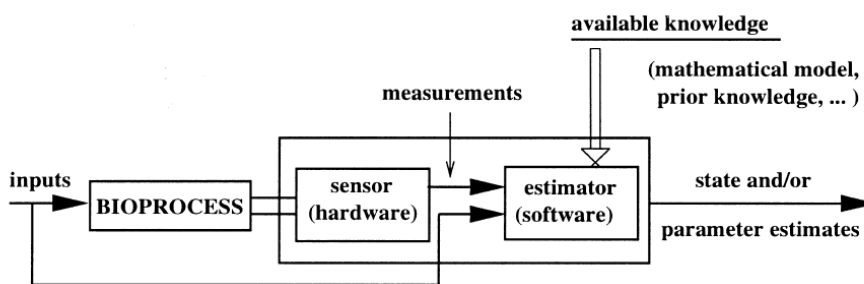


Figure 7: Principle of a soft-sensor (adapted from Farza, Busawon and Hammouri [82])

There are different approaches for soft-sensors. Either models are built basing on data-driven statistical methods like principal component regression (PCR), artificial neural networks (ANN) [80, 83-85] or recently also on independent component analysis and support vector machine [84]. Furthermore, they are approaches based on fundamental principles like chemical, biological or physical parameters and their impact on process variables. Latter can be based on first principles as for example growth stoichiometry, which has been proven to be effective [84, 86]. Out of growth stoichiometry mass balances are established. These equations are used to calculate missing variables [80]. For historical driven soft-sensors process data is needed to develop the soft-sensor tool. The soft-sensor has to be trained with previous data to create a connection between input- and output-variables. Especially ANNs work by connecting input and output via neurons (nodes) [83]. In contrast soft-sensors basing on first-principle models which illustrate the chemical, biological and physical relationships between quality and process variables require a detailed a priori knowledge of the process [84].

Most commonly soft-sensors are applied to predict the actual biomass concentration, as biomass measurements are difficulty accessible in real-time [86-88]. Conventional measurement

of biomass is a gravimetric determination of dry cell weight. It is defined as the insoluble fraction of mass in biomass which is dried [86].

1.6.2 Hard-type sensor vs. soft-sensors

In contrast to soft-sensors, hard-type sensors are quantifying biomass concentration based on optical methods like near infrared [89] or NADH, or via capacitance measurement [86]. Hard-type sensors have an advantage compared to soft-sensors. The latter are limited due to deviant culture conditions because missing values cannot be estimated accurately [89]. Capacitance measurements distinguishing viable and not viable cells, as only living cells are functioning as capacitor [86, 90]. Hence a hard-type sensor using a capacitance probe promises to be appropriate for a phase of cell-lysis. The amount of living cells by a hard-type sensor is not affected by changing culture conditions and therefor renders a control of specific substrate uptake rate in lysis phase possible. By controlling the q_s of glucose, a high cell growth/cell division can be maintained, which is necessary for protein E mediated lysis [69].

The dielectric permittivity of microbial dispersions which is directly proportional to capacitance showed to be linearly related to the membrane-enclosed volume fraction [91]. The more cells are present, the higher is the amount of spherical charged capacitors, which leads to enhanced capacitance in the suspension [92]. Hence only viable (non-lysed) cells are detected, and biomass concentration can be determined.

The electrical properties of cells like permittivity and conductivity are strongly dependent on the measurement frequency [93]. At low radio frequencies (0.1-10 MHz) the membrane becomes polarized by the so-called β -dispersion effect [94]. The suitable frequency of measurements depends on cell size, morphology and type [95]. The relative permittivity is estimated using two frequencies, a high frequency accounting for non-cellular background (10 MHz) and further a low frequency accounting for the permittivity attributed to living cells (1 MHz). The difference between the obtained values yields the relative permittivity $\Delta\epsilon$ [96].

1.6.1 Process control using a soft-sensor

The concept of a first principle soft-sensor is shown in Figure 8.

Substrate and biomass stoichiometry has to be investigated before the application of the soft-sensor. Additionally feed concentration as well as density of feed and base has to be known. These constants with bioprocess data together are used by a volume calculation tool, to supply the soft-sensor with the actual reactor volume.

Volume calculation is based on the mass balance, whereas the change of volume is calculated of mass in-flows minus out-flows (see Equation 1). Hence in a fed-batch changes of mass depend on substrate and base in-flow, as well as oxygen added and outgoing CO₂. (see Equation 2).

$$\dot{M}_{in} - \dot{M}_{out} = \frac{dM}{dt}$$

Equation 1: Mass balance

$$\frac{dM}{dt} = F_{s,in} + F_{b,in} + r_{O_2} * 32 - (r_{CO_2} * 44 + \frac{F_{a,out} * ex_{H_2O,out} * 60 * 18}{V_m})$$

Equation 2: Mass change in bioreactor

When dividing the mass change by the density of the broth, the value of the changing volume in the bioreactor is received:

$$\frac{dV}{dt} = \frac{(\frac{dM}{dt})}{\rho_v}$$

Equation 3: Volume change in bioreactor

The information about volume change is transferred to the soft-sensor which uses the information to estimate the current biomass in the fermenter. When specific substrate uptake rate or growth rate of the culture are controlled by the soft-sensor, biomass estimation is used to calculate the appropriate flow rate, which is regulated *via* PI flow controllers.

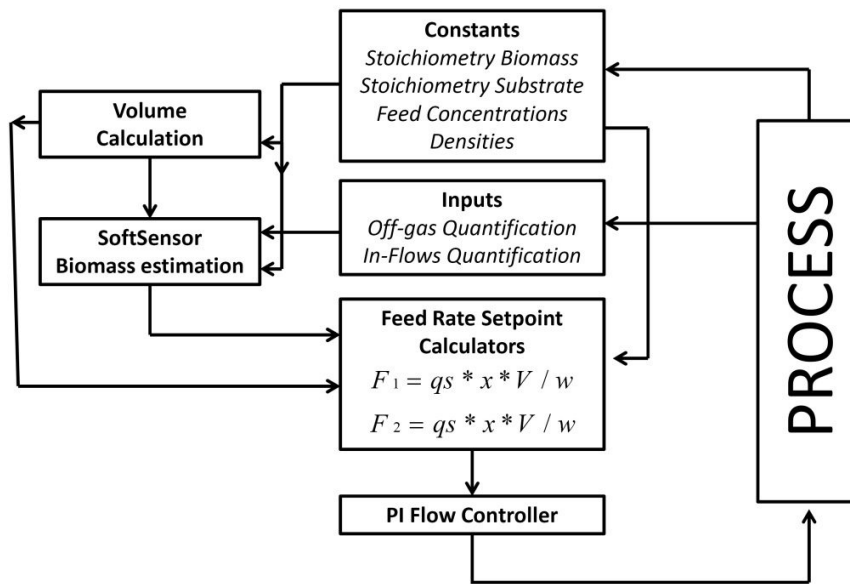


Figure 8: Outline of the soft-sensor concept (adapted from Sagmeister et al. [57])

For the control of substrate feed, the relation between substrate consumption and biomass growth, namely the biomass yield coefficient, needs to be known. As this yield coefficient is variable during recombinant protein production, an algorithm is used to deal with variability in the model [86, 97]. For example extended Kalman filters are applied to calculate complex non-linear algorithms [82, 98, 99]. But also a partial least squares (PLS) regression can be used to estimate the biomass yield coefficient.

Target variables are reconstructed basing on non-linear first-principle models of the process and current online measurements [98]. By utilization of the actual values for biomass concentration, substrate concentration, mass or volume of fermentation broth as well as the growth rate, a substrate feed rate is calculated, which is necessary to keep the process at a determined specific growth rate profile [97].

Sagmeister et al. [57] showed, that the soft-sensor assisted control of specific substrate uptake rates, namely specific uptake rates for D-glucose and L-arabinose, in a mixed-feed system is possible [57]. The specific uptake rate of L-arabinose is expected to be distinct for the product

titer in a pBAD expression system. This stands in contrast to target protein production control *via* total specific substrate uptake rate, which has been shown by Wechselberger et al. [100]. There variations in time-space yield have been sufficiently described by the total substrate feed rate, whereas the culture has been induced by a not metabolized sugar. Hence the inducer was added once and then the product titer depended on the feeding profile. The time-space yield is described as the relation of product to space and time, which means the amount of enzyme produced per culture volume over a defined time-span during fermentation process [100].

Due to the ability of soft-sensors to dynamically control process parameters like specific substrate uptake rate, they are a useful tool in process development. They offer the possibility to examine multiple levels of process parameters within one experiment and thereby advance process development [17]. In contrast to this application, where process knowledge is not always available, soft-sensors are a promising tool in manufacturing processes or pharmaceutical production processes where high amount of process data is gathered and soft-sensors facilitate to assure quality through feed-back regulation of the process, whereby deviations caused by distortions are considered. This is not possible when a process is controlled feed forward [85, 97]. Hence reproducibility of fermentations and process quality are enhanced. This goes in line with the PAT approach of the FDA [85]. Furthermore the high reproducibility of soft-sensors makes them a future tool for the control of fermentations in DoEs, because runs become more comparable.

1.7 Multivariate data analysis

1.7.1 Multivariate data analysis in bioprocess technology

During fermentations processes a high amount of data is collected. The measurements are necessary to monitor deviation in the process, which can lead to reduced productivities or product quality issues. The received data needs to be analyzed and interpreted to gain further knowledge about the process and the interaction between multiple parameters [101]. Statistical methods can be applied to describe connections between variables, but also to better understand the physiology of the microorganisms and for the development of prediction

techniques. As described in section 1.6, statistical methods like ANN and PLS are used for the application of soft-sensors to control processes and to do predictions.

Process understanding and process control are distinct to ensure high product quality in pharmaceutical processes [102] and it goes in line with the PAT approach of the FDA [85]. Furthermore, knowledge gained by statistical analysis can enhance process robustness [103].

Because of its high dimensionality and the co-linearity between multiple variables, the treatment of bioprocess data is very challenging [103]. Methods like PCAs are applied to reduce the dimensionality of data. In this approach original data from multiple fermentations can be transformed into a matrix of mutually uncorrelated variables, called principal components (PCs). These PCs are chosen to explain the highest possible variance in the data. The connection of original variables to PCs is defined by loading vectors. There are as many PCs chosen, that a predefined amount of variance in the data is explained [102].

PCAs have been used for process monitoring [102, 104], where a model was developed using historical data. This model later was then applied on online data [102].

Due to its variable reducing function, PCA is often used for the examination of data, whereas regressions are applied for further investigations [105]. MLR is used to examine connections between two and more variables and to describe one variable through alterations in other parameters. However, colinearities of variables used for MLR impede the interpretation of model coefficients. Principal component regression (PCR) is a combination of PCA and linear regression. Its advantage in comparison to MLR is, that the PCs are mutually orthogonal, which avoids the problem of collinearity between variables [105].

1.7.2 Rates and yield coefficients as physiological parameters

Variances of CQAs in a process are linked to CPPs by cell physiology, which can be described by calculating rates and yield coefficient concerning substrate and oxygen uptake as well as biomass, CO₂ and product production [106].

In performing a DoE investigating the impact on CPPs on CQAs, a knowledge space is built. When multivariate methods like MLR are applied to quantify the impact of variability of CPPs on

CQAs, the optimal conditions for the process are specified and a design space created. Within this design space variations of CPPs have no significant impact on CQAs [38].

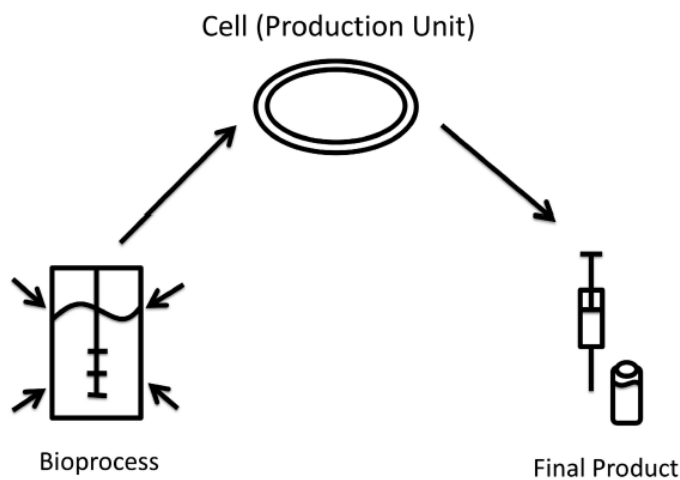


Figure 9: Changes in bioprocess parameters influence product quality via deviations in the cell as protein factory (adapted from Sagmeister, Wechselberger and Herwig [106])

But by only connecting CPPs and CQAs *via* statistical methods, the biological system in-between is unknown, seeming like a black box. To investigate in biological/physiological changes dependent on the variation of CPPs, these factors are linked to rates and yield coefficients concerning physiological processes of the cell (see Figure 9). Therefore multivariate methods are applied as well. An MLR explains the variance in yield coefficients and rates and thereby alterations in the organism caused by varying CPPs [106]. However, this approach can only be applied on variables that are not co-linear, e.g. on data obtained from DoEs.

To further investigate the system, yield coefficients and rates are related to CQAs, either by MLRs to explore the changes in CQAs dependent on the variability of yield coefficients and rates. When an MLR is no more appropriate because of too many input variables a PCA can be conducted to extract the main factors influencing variability in CQAs, and how these factors are linked to input variables.

1.7.3 Interpreting yield coefficients and specific rates

Yield coefficients and specific rates calculated out of process data inform about conversion of carbon into biomass, product, extracellular protein and CO₂. In a non-induced culture were no

by-product is produced, the whole carbon fed is found in biomass and CO₂ produced, which is illustrated in biomass and CO₂ yield coefficients, namely $Y_{x/s}$ and $Y_{CO_2/s}$.

After induction, the cell compensates additional energy required for recombinant protein by reducing biomass production. Hence the yield coefficient of biomass out of substrate ($Y_{x/s}$) decreases as well as growth rate decrease [29]. The smaller $Y_{x/s}$ due to metabolic burden in the cell imposes an enhanced $Y_{CO_2/s}$, as energy is needed for the plasmids. Beside $Y_{x/s}$ and $Y_{CO_2/s}$ also the product yield coefficient can be negatively affected by the metabolic burden [32]

When comparing $Y_{x/s}$ of two not induced cultures, whereas one bears a plasmid for recombinant protein production and one does not, the one bearing the plasmid has a smaller specific growth rate [31, 32]. Another study showed a connection with smaller specific D-glucose uptake rate [107].

By-product formation like acetate production under oxygen limitation triggers lower biomass and recombinant protein production, hence decreased $Y_{x/s}$ and decreased $Y_{p/s}$ [41].

Apart from recombinant protein formation, a CQA in this study supposed to closely be linked to changes in yield coefficients and rates is the lysis efficiency. The ability to E-lyse has been shown to depend on the physiological state of the cells [6, 7].

1.8 Aims of this study

This study aims to combine the BG platform technology to the pBAD system for recombinant protein production. Thereby the downstream process should be facilitated, as cell rupture methods like homogenization are no longer needed. As a result the downstream process would be shorter and less expensive, due to decreased equipment cost and further efforts like clean in place (CIP) and steam in place (SIP) which are linked to an additional process step.

The advantages of the pBAD system like tuneability of recombinant protein production offer the possibility to avoid high metabolic loads due to the production process, and to thereby keep the cells in an E-lysis capable state. To ensure an effective triggering of recombinant protein production, specific substrate uptake rate needs to be controlled. Through the application of a soft-sensor in process development, the accuracy of design space investigation should be raised,

as the CPPs are controlled during fermentation. This soft-sensor assisted control is in addition independent of spontaneous deviations during fermentation processes.

The goals of the study in bullet points:

- 1) Investigation of rhBMP-2 formation kinetics and process optimization.
- 2) Demonstration of the applicability of the BG platform technology for the release of recombinant proteins in a fed-batch process.
- 3) Development of understanding of the interaction between process parameters ($q_{s_{ara}}$ and $q_{s_{gluc}}$) and the lysis competence of the cellular population (boundaries for the applicability of the BG platform technology).
- 4) Demonstration of the applicability of first principle soft-sensor control strategies within process development along QbD principles.

2. Materials and Methods

2.1 *Upstream process*

2.1.1 *The strain*

An *E. coli* C41 strain containing two plasmids was used. The plasmids were coding for recombinant hBMP2 as well as the phage protein E. The genes for recombinant protein production were on the plasmid pBK-BMP, which originates from pBAD24 (provided by BIRD-C, Vienna, Austria). The system was inducible by L-arabinose whereas the strain is capable of metabolizing the inducer.

The gene E cassette for temperature inducible lysis (42°C) was on pGLysivb (provided by BIRD-C, Vienna, Austria).

2.1.2 *Media*

Media were used as described in DeLisa et al. [108]. For fed-batch and lysis phase the feed medium was applied with D-glucose as only substrate. During induction phase D-glucose and L-

arabinose were added according to the intended ratio of q_s glucose and q_s arabinose, as described in section 3.1. The overall sugar concentration in the feed medium was 400g/l, for the batch medium 20g/l. The batch medium was complemented with 20mg/l gentamycin and 20mg/l kanamycin.

2.1.3 Bioreactor setup

Multiple fermentations were conducted in a Techfors-S bioreactor (Infors, Bottmingen, Switzerland) of 10l maximum working volume (see *Figure 10*). This reactor was stirred with a 50 to 1500rpm “disk” type impeller. Gas was added by passing through a 0.2 μ m membrane-type filter. It was dispensed by a ring sparger. To measure oxygen and air gas flow rates, a thermal mass flow controller (Vögtlin Instruments, Aesch, Switzerland) was used.

Temperature was controlled by a deep-profile double jacket which was supplied by an external chiller, as well as an electrically powered heat exchanger. Sampling and harvesting was done with a sterilizable valve on the bottom of the fermenter.

For feed and base addition peristaltic pumps integrated into the Techfors-S bioreactor were used. Feed and base bottles were linked to the bioreactor *via* silicon tubing and quick connect couplings (Roth, Dautphetal, Germany) for sterile connections. The feed was added from the top of the bioreactor, whereas for the base an immersion tube (Infors, Bottmingen, Switzerland) was used.

The control of multiple parameters like pH, temperature, O_2 and air inlet, pressure, agitation speed and DO_2 in the Techfors-S bioreactor was possible by a touch screen control panel connected to the bioreactor.

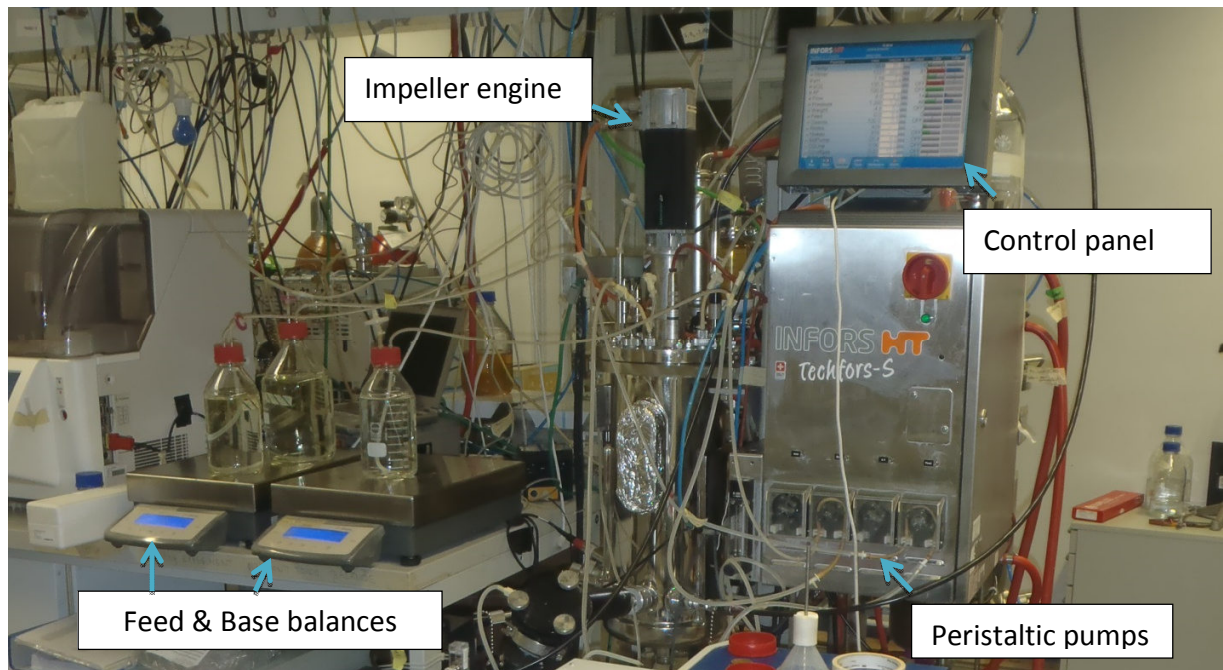


Figure 10: Bioreactor setup

2.1.4 Bioreactor instrumentation

Due to gravimetrically quantification of substrate and base applied, base and feed solutions were placed on balances (Sartorius, Göttingen, Germany) to record the weight loss.

The ports on the top plate of the fermenter were used for pH probe (Hamilton, Reno, USA), pressure sensor (Keller, Winterthur, Switzerland), overpressure valve (Infors, Bottmingen, Switzerland), dissolved oxygen probe (Hamilton, Reno, USA) and a septum. Additionally an annular type permittivity probe (Aber Instruments, Abersystwyth, Wales, UK) was applied to measure relative permittivity and conductivity of the fermentation broth.

CO₂ and O₂ concentrations in the off-gas were quantified by a gas analyzer (M. Müller AG, Esslingen, Switzerland) whereby CO₂ was analyzed based on NDIR and O₂ on a paramagnetic principle.

All signals were recorded by the process management system Lucullus (Secure Cell, Schlieren, Switzerland), which was as well used to control the process. Volume estimation tools and the soft-sensor were implemented in Lucullus integrated interfacing tool “Sim-Fit”. The soft-sensor applied here (outlined in section 1.6.1) was used to estimate biomass concentration.

2.1.5 Culture mode

Product stocks were stored on -80°C. The fermenter including 4 l Batch medium was sterilized for 20min at 121°C. The preculture was grown for 8 hours in shake flasks at 35°C. The bioreactor was inoculated with 200ml preculture of an OD of 1-2. Cultivation parameters are described in Table 1.

Table 1: Parameter set-points in the bioreactor

Parameter	Set-point
Stirrer speed	1450 rpm
pH	7.2
Pressure	1.2 bar gauge
DO2	>30%
Temperature	35°C (42°C during lysis phase)
Gas in-flow rate	1.5 vvm

After the end of batch phase, which was detected by a decline of CO₂ off-gas signal, a fed-batch using a feed of 400g/l D-glucose was started. A forward-controlled exponential ramp according to Equation 4 with a specific growth rate of 0.2h⁻¹ was run. The calculation of the initial feed rate was conducted as shown in Equation 5. The initial biomass concentration to calculate the feed rate was obtained via OD600 measurement according to Equation 6. The measurement was conducted with a photometer (Genesys 20; Thermo Fisher, Waltham, Massachusetts, USA).

$$F_{(t)} = F_0 * e^{\mu * t}$$

Equation 4: Calculation of the exponential feed rate

$$F_0 = \frac{\mu * c_x * M_s * \rho_{feed}}{x_s * Y_{x/s} * M_x}$$

Equation 5: Calculation of the initial feed rate

$$c_x = OD_{600} * 0.46$$

Equation 6: Biomass concentration calculation

The constant 0.46 had been empirically defined.

Parallel to the beginning of fed-batch, the soft-sensor was initialized with the estimated biomass concentration estimated *via* OD correlation inserted as starting value. When the biomass concentration reached a value of 15g/l due to soft-sensor estimation, glucose feed was stopped and the culture was induced *via* an L-arabinose pulse of 25ml of 400g/l L-arabinose solution to a final L-arabinose concentration of 2.5g/l. When the off-gas CO₂ signal after the pulse decreased and thereby indicated substrate limitation, a mixed substrate exponential feed was started.

In this fed-batch phase the feed was feedback controlled by the soft-sensor, meaning that the soft-sensor calculated a feed rate on the basis of actual estimated biomass concentration and the q_s set-point. Equation 7 shows the feed rate calculation.

$$F_{(t)} = \frac{q_{s_total} * c_x * V * \delta_{feed}}{c_s}$$

Equation 7: Feed rate calculation using soft-sensor estimated values to control q_s

In order to obtain comparable signal to noise ratios in course of quantitative bioprocess evaluation from off-line data [109] sampling intervals were chosen according to Equation 8.

$$Window = \frac{SNR * Error}{67 * \mu}$$

Equation 8: Calculation of the sampling window during the fermentation process

The desired signal to noise ratio was 24 and the expected error on the biomass estimation 1%.

Approximately 10ml of sample were taken when the estimated biomass values reached 19g/l, 24g/l, 29g/l and 35g/l.

When the estimated biomass concentration reached a value of 35g/l, feed was changed back to glucose feed. After adding polypropylene glycol to a final concentration of 0.5 g/l to prevent foam formation, the reactor temperature was shifted to 42°C, to induce protein E formation.

During lysis phase samples were taken every 30min. The feed was controlled via the capacitance signal, and was set to hold a q_s of 0.8g/g/h. The process was stopped after 180 minutes lysis phase.

During and after the process, the measurements showed in Table 1 were conducted.

Table 2: Online and offline measurements of the process

	Quantified analyst	Analytical method
Online	Air flow in	Mass flow controller
	Oxygen flow in	Mass flow controller
	CO ₂ in the off-gas	NDIR
	O ₂ in the off-gas	Paramagnetic sensor
	Base added	Balance
	Feed added	Balance
Offline	Extracellular protein	BCA
	Biomass produced	Dry cell weight determination
	Molarity of base	Titration
	Feed concentration	Ion exchange HPLC
	Product produced	Reversed phase HPLC
	By-product formation	Ion exchange HPLC

2.2 Off-line analytics

2.2.1 Biomass dry cell weight

DCW determination was done in duplicate. In an eprouvette 2ml sample was centrifuged at 4332RZB, 10min, 4°C. The pellet was washed, centrifuged again and then dried for 48 hours at 105°C. The DCW was achieved gravimetrically.

2.2.2 Extracellular protein determination

The supernatant of the centrifugation described in section 2.2.1 was stored at -20°C. Total extracellular protein in the supernatant was quantified photometrically (bicinchoninic acid method) following TCA precipitation, as described in detail in the appendix (sections 6.1.1 and 6.1.2). Total extracellular protein was determined for all samples taken before lysis phase.

2.2.3 Homogenization and cell rupture

A duplicate of 2ml fermentation broth was centrifuged (4332RZB, 10min, 4°C), washed once and stored at -20°C for further use. The supernatant was stored separately at 20°C. To prepare samples for homogenization, the pellets were resuspended to 20ml in a buffer (50mM Tris, 5mM DTT, 1mM EDTA, pH 8). All samples were suspended in the buffer, but only samples taken before lysis phase were homogenized (Avestin EmulsiFlex©, Ottawa, Canada) in a continuous mode at 1500bar for 3 passages.

2.2.4 Purity determination of rhBMP-2 via SDS gel analysis

1ml of the homogenized respectively the lysed sample in the homogenization buffer was centrifuged (16600RZB, 10min, 4°C). For homogenized samples, the pellet and the supernatant were separately reduced in buffer described in Laemmli (1970) [110] on the heating block (5min, 95°C). Pellets of lysed samples and the supernatant of the centrifugation described in section 2.2.1 were treated the same way. 30µl of each sample, as well as 10µl standard (SeaBlue® Plus2 Pre-Stained; Invitrogen, Carlsbad, California, USA) were loaded onto sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS) gels (8-16%; GE Healthcare, Chalfont St. Giles, Great Britain).

After running the gels (160V, 60min) they were stained overnight in a sensitive Coomassie solution (0,02% (w/v) Coomassie Brilliant Blue G 250, 5% (w/v) Aluminium Sulfat- (14-18)-Hydrat, 10% (v/v) Ethanol, 2% (v/v) Orthophosphoretic Acid, distilled water). Stained gels were scanned and analyzed using the software Image Lab (Bio-Rad, Hercules, California, USA). To investigate the purity of hBMP2, before determined reference bands as well as a band at 13kDa for the monomer of hBMP2 were marked and the intensity of bands compared. Each sample was evaluated in triplicate on three gels.

2.2.5 hBMP-2 titer investigation via RP-HPLC

As described in section 2.2.3, 1ml of the homogenized respectively lysed samples in homogenization buffer was centrifuged (16600RZB, 10min, 4°C), the supernatant discarded and the pellet resuspended in solubilization buffer (6M GuanidinHCl, 50mM β-Mercaptoethanol,

10mM Tris, 10mM iodacetamide, pH 7.6). The volume of buffer depends on the hBMP-2 concentration. The suspension should not exceed 5g/l hBMP2. For solubilization, samples were shaken overnight in 15-ml plastic tubes (Greiner Bio-One International AG, Kremsmünster, Austria).

Then samples were transferred to 1.5ml reaction tube (Eppendorf, Hamburg, Germany) and centrifuged (14000rpm, 30min). The supernatant was analyzed on HPLC (UltiMate 3000; Thermo Fisher, Waltham, Massachusetts, USA) with a reversed phase column (RP-1S, Thermofisher, Waltham, Massachusetts, USA) using an in-house developed method (flowrate 0.5ml/min; eluent: 95% acetonitrile, 5% H₂O, 0.1% Trifluoric acid; linear gradient 0-100% eluent in 10min, 30°C, UV detector).

2.2.6 Flow cytometry: Investigation of lysis efficiency

To detect the advance in protein E mediated lysis, flow cytometry was done, wherefore a Cube 6 system (Partec, Münster, Germany) was used. Dilution of the sample was chosen to show maximal 4,000 particle counts per second, when the sample speed was 1 μ l/s. The machine had a 488 nm blue solid state laser, whereas five optical parameters were available. These were three fluorescence channels as well as forward and side scatters (FSC/SSC). The wavelength detected by the fluorescence channels (FL) were 527nm for FL1 (band-pass filter, bandwidth 30 nm), 590nm for FL2 (bandpass filter, bandwidth 50nm) and 630nm for FL3 (longpass filter).

The method applied was adapted from Langemann et al. (2010) [7]. Diluted samples were dyed using the fluorescent dyes RH 414 (red, abs./em.: 532/718nm) and DiBAC4(3) (green, abs./em.:493/516nm) (both: AnaSpec, Fremont CA, USA). RH 414 unspecifically stains plasma membranes, whereas DiBAC4(3) stains cells only with a disrupted membrane potential.

The channel FL2 was used to separate cells from noise. Only cells were further investigated. The FL1 signal discriminated living from not living cells, due to DiBAC4(3) staining. To distinguish lysed (empty) cells (bacterial ghosts) from full cells, the FSC was used, because the FSC signal for full cells was higher than for empty ones. In a density plot with FSC as abscissa and FL1 as ordinate, gates (R1-R3) were defined to separate three types of cells (see Figure 11):

- R1 - high FSC signal ('full') and low FL1 signal (viable): whole live cell
- R2 - high FSC signal ('full') and high FL1 signal (dead): whole dead cell
- R3 - low FSC signal ('empty') and high FL1 signal (dead): Bacterial Ghost

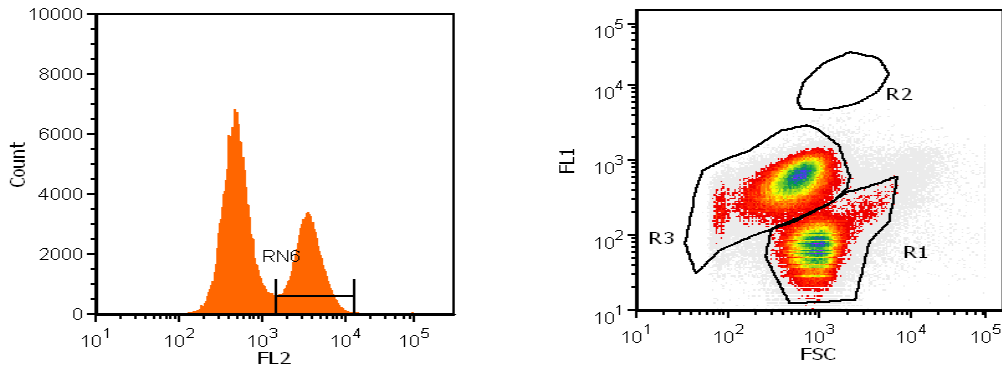


Figure 11: Measurement of lysis efficiency. Discriminating cells from other particles using FL2 (left picture) and distinction of bacterial ghosts, living cells and dead cells with FSC and FL1 distinguish (right picture).

Lysis efficiency (LE), hence the percentage of bacterial ghosts in all cells, was to be determined as described in Equation 9.

$$LE = \frac{R3}{(R1 + R2 + R3)} * 100\%$$

Equation 9: Calculation of lysis efficiency basing on the cell counts in three different gates

2.3 Multivariate data analysis

Online as well as offline measured data was collected in Lucullus (see Table 2 for overview of offline and online measurements). For treatment, data was exported and further rates, yields and specific rates were calculated using a Matlab tool (equations were adapted from Wechselberger et al[86] and Sagmeister et al. [106]. For multivariate data analysis, an information matrix was created, in which the over-time data of multiple fermentations was reduced to one value per variable and fermentation, as described in detail in the results section 3.2.4. Hence the dimension time was no longer needed (unfolding, one observation per fermentation run). The dimensions in the unfolded matrix were multiple fermentations and

variables like the mean of yield coefficients and balances as well as the end-values of titers and the lysis efficiency. The flow of information processing is shown in Figure 12.

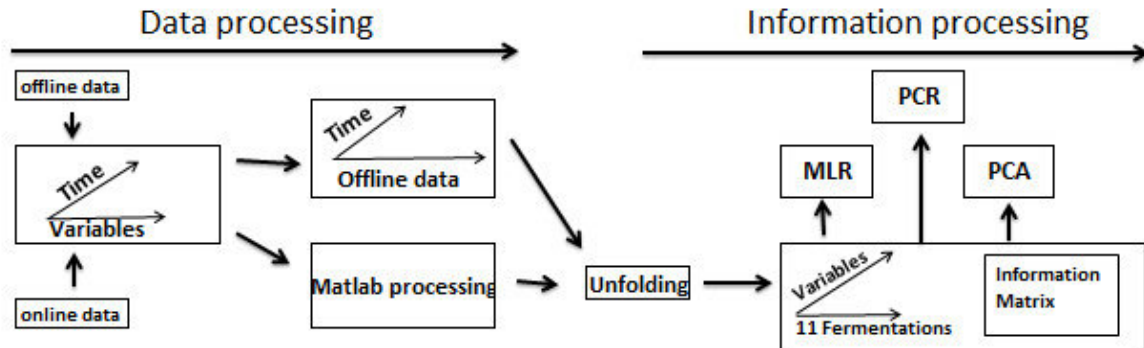


Figure 12: Work flow of data and information processing.

Table 3: Variables used for the PCA and how the unfolded values were calculated

Variable(s)	Origin of value	Variable	Origin of value
$q_{S_{ara}}, q_{S_{gluc}}, q_{S_{total}}$ [g/g/h]	Mean over induction	Purity BL/AL [%]	Purity at start/end of lysis phase
xAra [%]	% of arabinose in feed	YTS BL/AL [g/l/h]	YTS before/ after lysis
Induction period [h]	Duration of mixed-feed phase	Titer extracellular protein [g/l]	Last sample of induction phase
μ [1/h]	Mean over induction	Titer BL/AL [g]	Titer start/end lysis
$Y_{x/s}, Y_{CO_2/s}$ [c-mol/c-mol]	Mean over induction	Lysis efficiency [%]	Sample with highest lysis efficiency
$Y_{O_2/s}$ [mol-c-mol]	Mean over induction	Lysis efficiency BL [%]	Lysis competent cells at lysis induction
RQ [c-mol/mol]	Mean over induction	Reaction constant	Coefficient to describe lysis kinetics
rhBMP-2 titer [g/l], [g/g], [g]	Last sample in induction phase	rhBMP-2 recovery [%]	% of rhBMP-2 found after lysis compared to titer at lysis start
qp [g/g/h]	Mean over induction		

3. Results and Discussion

3.1 Process Description and DoE Design

Figure 13 shows the single steps for the definition of platform constrains, starting with a risk assessment, where process parameters ($q_{s_{ara}}$ and $q_{s_{gluc}}$) were assessed to be critical (critical process parameters, CPPs). These process parameters were assessed to be critical in a previous study [69]. In this master thesis a DoE was done to create knowledge about rhBMP-2 titer and lysis efficiency among variable $q_{s_{ara}}$ and $q_{s_{gluc}}$ space.

Knowledge about physiological constrains (onset of catabolite repression) of the system was based on the results of Sagmeister et al. [57], who had conducted pulse experiments to investigate the limits of the mixed feed pBAD system in concern of catabolite repression (maximum uptake of L-arabinose ($q_{s_{ara}}$) in the presence of uptake of D-glucose ($q_{s_{gluc}}$), as outlined in section 1.3.3.

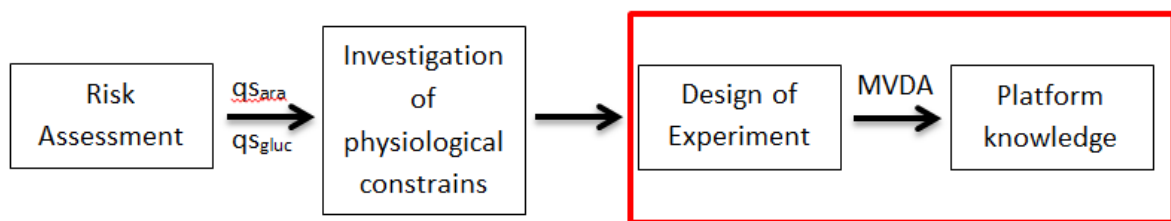


Figure 13: The work flow from risk assessment to platform knowledge. The scope of this work is marked in a red frame.

Based on this knowledge, an orthogonal experimental design was set inside these limits to create a knowledge space for the platform system. Set-points for specific substrate uptake rates for both D-glucose and L-arabinose were set to be 0.05, 0.125 and 0.2 g/g/h. Eleven fermentations were planned including all possible combinations of these three q_s levels (two level full factorial CCF design) for both substrates. Additionally, the center point was conducted in triplicate to evaluate the experimental error. Hence the total specific substrate uptake rate varied between 0.1 and 0.4 g/g/h. The left picture of Figure 14 shows the points of the DoE, whereas each point represents a fermentation run. The order of runs was distinct to the numbering in Figure 14.

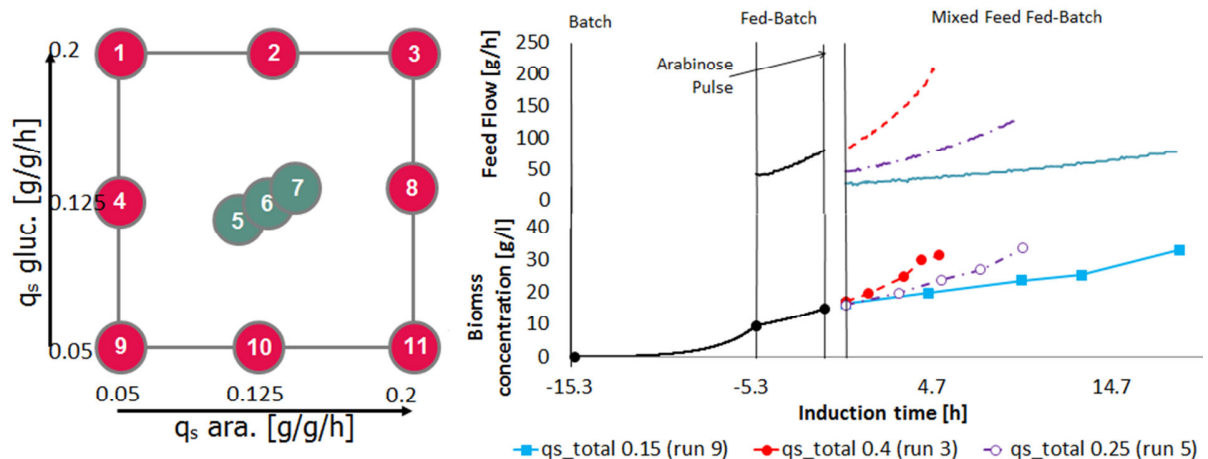


Figure 14: Orthogonal experimental design (left) and an overview over the feed flow and biomass concentrations during different phases of the process (right)

Process parameters in batch, fed-batch and lysis phase were the same for all eleven fermentations. The specific substrate uptake rate of D-glucose and L-arabinose in the “mixed feed fed-batch phase” were in focus of investigation (phase where recombinant protein is produced). To achieve a set of desired set points of q_s L-arabinose and q_s D-glucose, the ratio of D-glucose and L-arabinose in the feed was prepared respectively. The total specific substrate uptake rate was controlled by the soft-sensor. For fermentation 3, but also fermentations 5, 6, 7 and 9 the feed medium in induction phase was composed of 200g/l D-glucose and L-arabinose each, as the specific q_s of the two substrates had a ratio of 50:50. For fermentation 3 the total q_s was controlled at a value of 0.4g/g/h, whereas in fermentations 5 and 5 it was set to 0.25g/g/h.

Overall specific substrate uptake rates were set by the feeding rate. The right part of Figure 14 shows how the feeding rate differed for three runs with varying q_{s_tot} . Due to differing feeding strategies, biomass growth was distinct. Hence, duration of induction phase varied depending on the total specific substrate uptake rate. However, all processes were run to a final biomass concentration of ~ 35 g/l.

In contrast to the design showed in Figure 14, some points in the DoE were slightly changed. At the end of induction phase in fermentation 1 the pO_2 decreased beneath 30%. Hence the lysis efficiency was no more comparable and the fermentation had to be repeated.

The low-low point was not conducted at 0.05 g/g/h $q_{s_{gluc}}$ and 0.05 g/g/h $q_{s_{ara}}$ like it had been planned. But lysis efficiency for low total specific substrate uptake rate was expected to be low. Therefore the point had been chosen to be at 0.075 g/g/h $q_{s_{gluc}}$ and 0.075 g/g/h $q_{s_{ara}}$ instead. The third change to the original DoE design was that the point with $q_{s_{gluc}}$ 0.05 g/g/h and $q_{s_{ara}}$ 0.2 g/g/h was not conducted at all. Raised stress at high inducer concentrations were associated with low lysis efficiencies. Additionally L-arabinose is an expensive substrate compared to D-glucose and therefore utilization of that high amounts may not be rentable.

3.1.1 *Soft-sensor assisted process control in combination with a mixed-feed system*

The aim of applying a soft-sensor was to control the specific uptake rates of D-glucose and L-arabinose (parameters to be investigated within the multivariate study) on a defined level. This control of feeding rate based on estimated values of actual biomass concentration. Hence, the quality of feeding control depended on the accuracy of the estimated values.

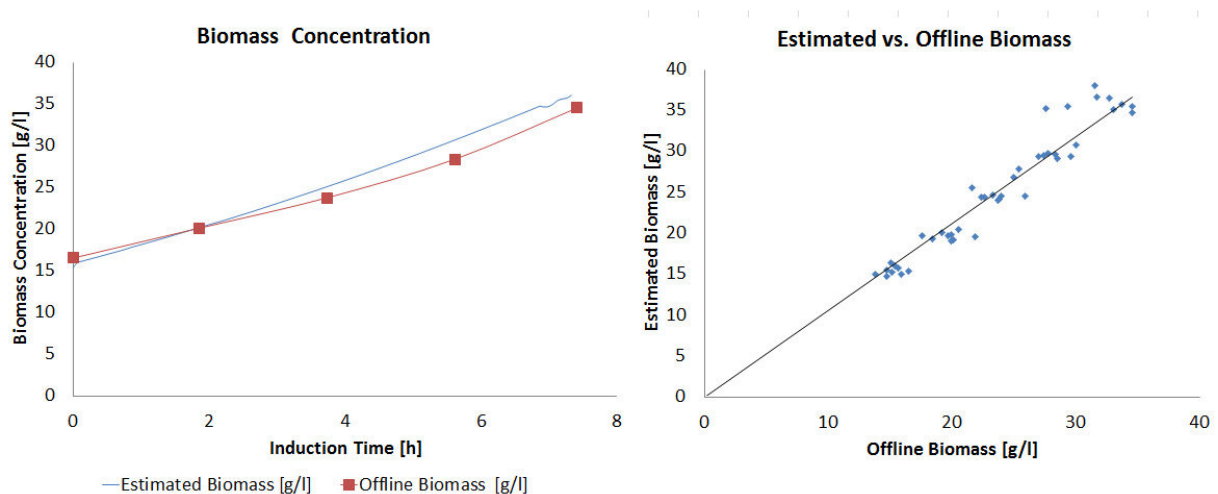


Figure 15: Estimated vs. offline measured biomass concentration over time in fermentation 8 (left) and the correlation between estimated and offline measured biomass (right)

The picture on the left in Figure 15 shows exemplarily the trend of estimated biomass and offline determined biomass over time for fermentation 8. A similar trend can be seen, when multiple offline biomass measurements of all fermentations are compared to estimates (see Figure 15, graph on the right). The real values were subtracted from the estimated values, to receive the residuals for a theoretical correlation with a slope of 1. For these residuals a 2-

sample t-test showed that during fed-batch and the first phase of induction (biomass DCW below 25g/l), estimated and offline determined biomass values coincided better than for the end phase of induction (see Figure 16 left side). Hence the H_0 “There is no difference in residuals between first and second phase of induction” was refused in favor of the alternative hypothesis stating a difference between these two phases. In latter period the soft-sensor tended to overestimate the actual biomass concentration. This led to an early termination of induction phase in some fermentations without reaching the determined value of 35g/l lysis induction. The box-plot in Figure 16 shows the distribution of residuals for samples with a DCW of smaller than 25g/l and for samples with higher biomass concentrations. The residuals vary less at lower biomass concentrations. Furthermore, the box-plots show the tendency to overestimate at higher biomass concentrations.

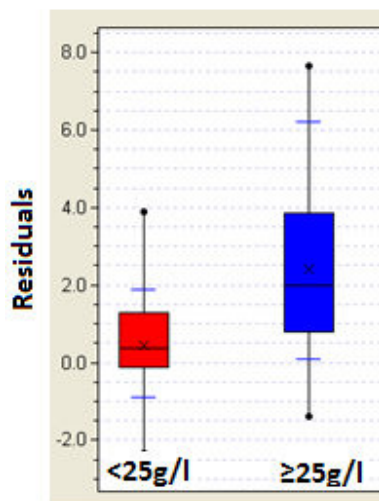


Table 4: Results for the 2-sample and a paired t-test

2-sample t-test		Paired t-test	
Mean	0.45 ± 1.3	Mean	32.7 ± 1.84
cx < 25g/l		DCW	
Mean	2.40 ± 2.40	Mean	35.7 ± 0.94
cx \geq 25g/l		estimated	
		cx	
p	0.0006	p	0.0009
α	0.05	α	0.05

Figure 16: Box-plot for the distribution of residuals in samples with low (left) and samples with higher (right) biomass concentration. The box of the box-plot is confined by the first and the third quartile. The line in the box represents the median and the x the mean. The whiskers mark 1.5 times of the interquartile distance.

The estimated biomass values of the last samples in induction phase differ significantly from the biomass DCW, as evaluated with a paired t-test (see

Table 4 on the right). Therefore the H_0 was formulated as “Estimated biomass values are the same or smaller than the biomass DCW”, and the alternative hypothesis as “The estimated biomass is higher than the biomass DCW”.

Fermentations 1a and 2 were stopped at DCW concentrations between 29 and 30g/l. Figure 16 compares offline measured biomass concentrations of the last sample in induction phase to

estimated values. Only fermentation 7 reached a biomass DCW of over 35g/l. Fermentations 4 and 8-10 had biomass DCW values between 33 and 34.5g/l (see biomass values at the end of fermentation in Table 6 in the appendix).

The average deviation between the set-point for total specific substrate uptake rate and its real value was 11%. 7% error was expected at the estimation of biomass. An important factor in biomass estimation was the feed concentration, which was wrongly stated in the soft-sensor. In the soft-sensor the set-point of feed concentration was stated, but in reality deviations from the planned feed concentration occurred. The real feed concentration was measured *via* RP-HPLC after the fermentation run. Hence, it was not considered for the soft sensor control strategy. The inaccuracies in feed-preparations were in average 7%. Table showing the errors of specific uptake rates for each fermentation are found in the appendix in Table 7.

During one induction phase the total q_s value varied 6% in average. The trend of specific substrate uptake rates for D-glucose and L-arabinose over induction phase is exemplarily shown for fermentation 8 (see left part of Figure 17).

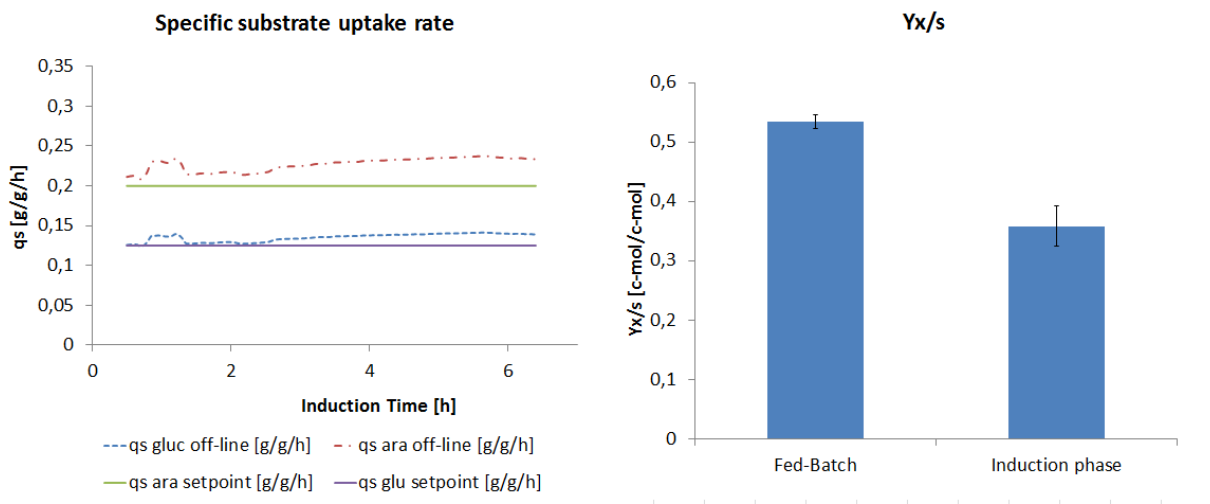


Figure 17: Comparison of set-points and real specific substrate uptake rates in fermentation 8 (left graph) and the means of offline measured $Y_{x/s}$ for fed-batch and induction phase of multiple fermentations (right)

The right picture in Figure 17 shows the average biomass yield coefficient $Y_{x/s}$ for fed-batch and induction phases. This yield coefficient was significantly higher in fed-batch phase than during protein production. Therefore, a 2-sample t-test with H_0 "The biomass yield coefficient during induction phase is higher or equal than the yield coefficient during fed-batch phase" and H_1

“The biomass yield coefficient during induction phase is smaller than the one in fed-batch phase” was conducted. At a confidence level of 0.05, the test was significant ($p > 0.0001$).

The biomass yield decreases after recombinant protein production due to higher energy demand of the cell for this unusual burden [29]. As outlined in section 1.6 the soft-sensor is based on the carbon and DoR balance. Hence its biomass estimations are independent of changing yield coefficients [17]. Therefore, no prior knowledge about the change of yield coefficients due to recombinant protein production needs to be known. This enables the application of the soft-sensor in process development.

The starting value of biomass investigated via OD600 measurement was not seen as a big source of error, as the estimated biomass values at the start of induction phase showed no high deviations from the biomass DCW. Additionally it is assumed that even though the initial value may be too high or too low, the biomass concentration levels off by adaption to the feeding rate. Trends of yield coefficients, rates and specific rates (as for example shown in Figure 23) support this assumption. In the first hours of induction phase the values vary more than at the end of induction phase.

Overall, the error in specific substrate uptake rate control was 11%. This is comparable to the dynamic q_s control of Sagmeister et al. [17], which had an error of 10% in average. Measurement of feed concentration before fermentation start would have reduced the error in process control.

Another source error in biomass estimation might be linked to extracellular protein production and acetate formation. This is because the applied soft-sensor did not consider by-product formation in its estimation of actual biomass. This assumption goes in line with the increasing error of estimation at increasing recombinant protein concentrations (see Figure 20). Extracellular protein concentration at the end of induction phase was as shown in Figure 18. However, no significant connection between extracellular protein formation and the difference between estimated and offline measured biomass concentration was found. In neither of the fermentations acetate was produced during induction phase.

Extracellular protein production is further outlined in section 3.2.2.

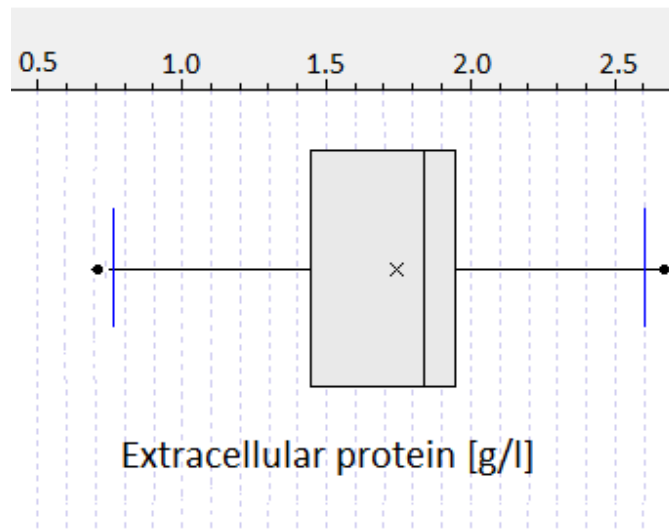


Figure 18: Measurements of extracellular protein concentrations at the end of fermentations (left). The box of the box-plot is confined by the first and the third quartile. The line in the box represents the median and the x the mean. The whiskers mark 1.5 times of the interquartile distance.

3.2 Data and information processing

Data was exported and – together with offline measured data – used in a Matlab tool for the calculation of yield coefficients, rates and specific rates (equations were adapted from [86, 106]).

As example for the richness of data, the complete data set of fermentation 4 is shown in Figure 19 (Online measurements), Figure 20 (Offline data) and Figure 23 (Processed data). Online data is shown over the whole process time.

3.2.1 Online data

Off-gas data (Figure 19), especially the CO₂ signal reflects the growth of the organism. The raise of the CO₂ signal until 6.1h process time reflects batch phase. The following four hours a fed-batch was run. After 9.2h fermentation, the culture was induced using 2.5g/l L-arabinose. The shift in the O₂ signal reflects the addition of the sugar, wherefore the reactor pressure had to be decreased for a minute to facilitate setting the pulse. The end of induced fed-batch (25.3h) was directly followed by the onset of lysis phase (25.5h). Again the pressure was set down for some seconds to add 0.5g/l PPG, which is reflected in a shift of oxygen in the off-gas.

After lysis induction, 1l/min of pure oxygen was added additionally to 5l/min air. Furthermore, the controlled specific substrate uptake rate was set to 0.8g/g/h. These, as well as the temperature shift to 42°C caused the sharp increase in CO₂ and O₂ outlet.

The permittivity signal (Figure 19) increased with raising biomass. After lysis induction (23.8h) the signal decreased due to progressive cell death (protein E mediated lysis). At the beginning of lysis phase the conductivity signal showed an opposed trend, since the content of the cytoplasm was released into the medium. But in the last phase of lysis, the conductivity signal decreased. This might be due to metabolism of released cell content by still living organisms and because of the high feeding rate, which led to dilution of the cell content with D-glucose feed. During batch-phase the conductivity signal was constant, and started to decrease in fed-batch phase.

The reducing of pressure for the application of the L-arabinose pulse or to inject antifoam was necessary. But the variation in pressure had an impact on oxygen saturation and pH. Figure 19 shows the trend of dissolved oxygen over fermentation time. When the L-arabinose pulse was done, the dissolved oxygen decreased due to pressure changes, but it was held over the critical value of 30%. This was not possible at lysis induction. There, the dissolved oxygen was transiently at zero, although pure oxygen was added (0.3 vvm) and the reactor was pressurized. The pH is sensitive to changes in pressure as this induces variations in the concentration of dissolved CO₂ as well as a changing feeding profile. The second is due to by-product formation and its catabolism. At the end of batch phase for example, acetate might be accumulated in the medium. When the whole substrate is consumed, the acetate is degraded by the bacteria, which induces a raise in pH. It was not possible to directly correct this shift in pH, because the pH was solely adjusted with NH₄OH, and no acid was employed. Hence, the pH decreased only by the consumption of substrate and thereby the increasing amount of CO₂ in the medium.

To control the pH on a constant level is important due to its critical function in the fermentation process. Depending on pH level protease activities differ and they have different effects on proteolytic degradation of recombinant protein [111, 112]. Temperature and pH are interacting parameters which influence growth rate. Only at non-stressing pH values (pH 6-9) no interaction of the two parameters was found concerning *E. coli* growth [113]. This is important in lysis phase, where a temperature shift from 35°C to 42°C may already stress the cells. The raise in

temperature could induce a stress response, where heat shock proteins are built. Furthermore, heat is associated with the denaturation of recombinant proteins [32, 114]. A shift in temperature is expected to have an almost instantaneous physical effect on the bacterial membrane and its composition [115] and the membrane plays a crucial role in E-lysis.

Taking the shock of the temperature shift, adaptation to higher feeding rates, the addition of PPG as well as the possible pH and dissolved oxygen changes by pressure decrease at the onset of E-lysis phase, this period of fermentation can be seen as crucial for the process concerning lysis efficiency. This is due to the sensitivity of the E-lysis ability on the physiological state of the cell [6, 7]. Additionally temperature shifts may cause plasmid loss [116].

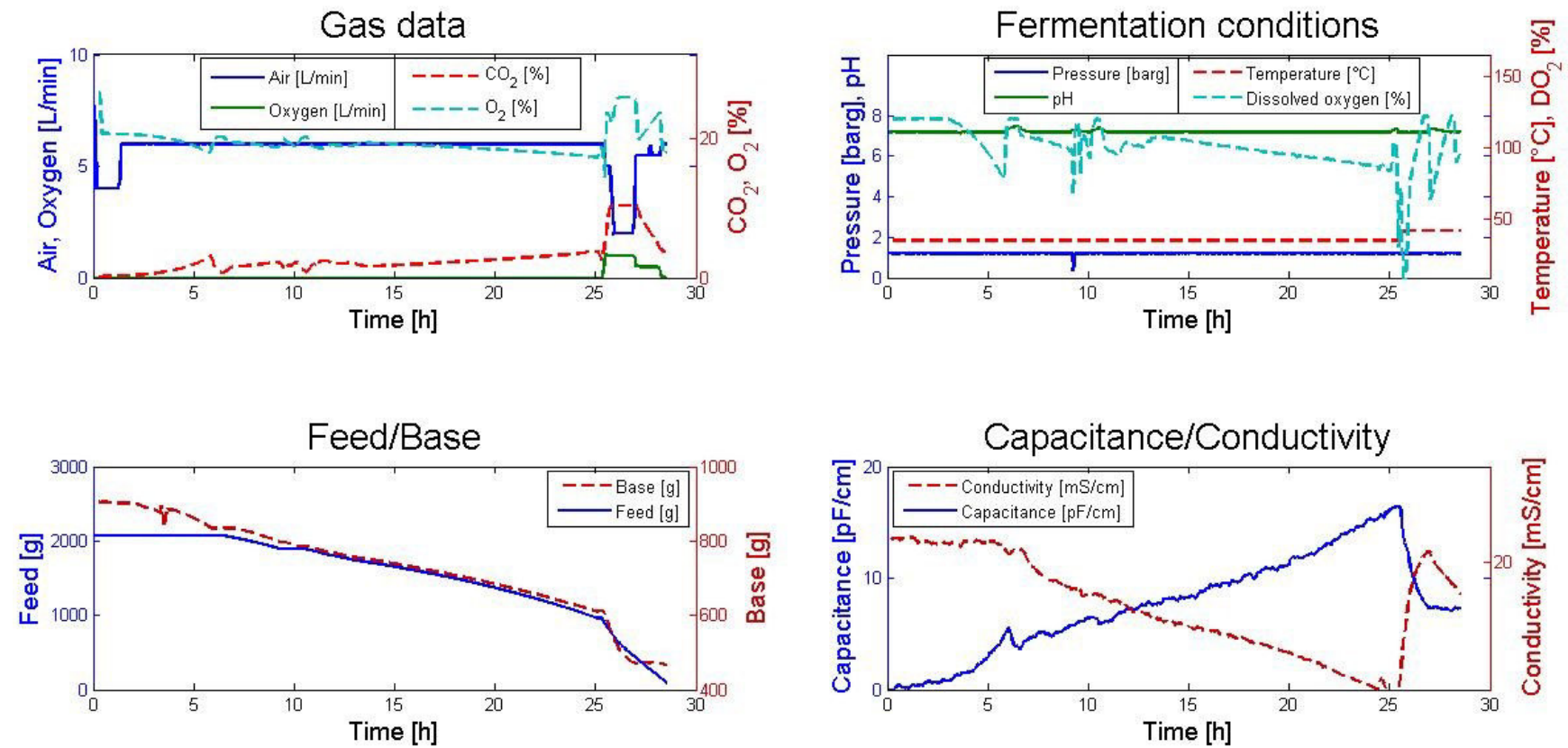


Figure 19: Online data of fermentation 4. Gas data like air inflow (blue), oxygen inflow (green), CO₂ (red) and O₂ (cyan) in the off-gas are shown in the graph on the top left. On the top right fermenter pressure (blue), pH (green), fermentation temperature (red) and dissolved oxygen (cyan) are seen. In the bottom graphs feed (blue) and base (red) balance weights on the left are presented as well as capacitance (blue) and conductivity (red) measurements on the right.

3.2.2 Offline data

Figure 20 shows the continuous increase of biomass concentration during induction phase. When the pulse was set (-1.0h) the biomass concentration was 15.9g/l. During induction phase it increased to a final concentration of 34.6g/l. The extracellular protein concentration in contrast was at the start more stable before it began to rise exponentially to a final concentration of 0.9g/l.

Lysis efficiency is defined as the percentage of bacterial ghosts in the whole cell population. It was investigated *via* flow cytometry (section 2.2.6). E-lysis competent cells are living cells, which are expected to be able to E-lyse (E-lysis competent cells are further outlined in section 3.1.3).

$$E - \text{lysisCompetentCells} = c_{x0} - c_{x0} * \left(\frac{LE_t}{100}\right) - (c_{x0} - c_{x0} * \left(\frac{LE_{final}}{100}\right))$$

Equation 10: E-lysis competent cells

As Equation 10 shows, E-lysis competent cells are calculated out of the biomass and lysis efficiency, whereas the cell population which is not capable to E-lyse is considered. From lysis onset on the concentration of E-lysis competent cells decreased until there were no E-lysis competent cells left anymore. The trend of lysis efficiency was opposed.

Product data is shown using cumulated substrate instead of time as abscissa. This was done to have a dimension in which multiple fermentations are commensurable, because due to varying q_s set-points between the induction phases of fermentations, the process time varied. But as all fermentations were grown to a biomass concentration of 35g/l, approximately the same amount of substrate was added. Product titer and product formations kinetics are discussed in section 3.2.

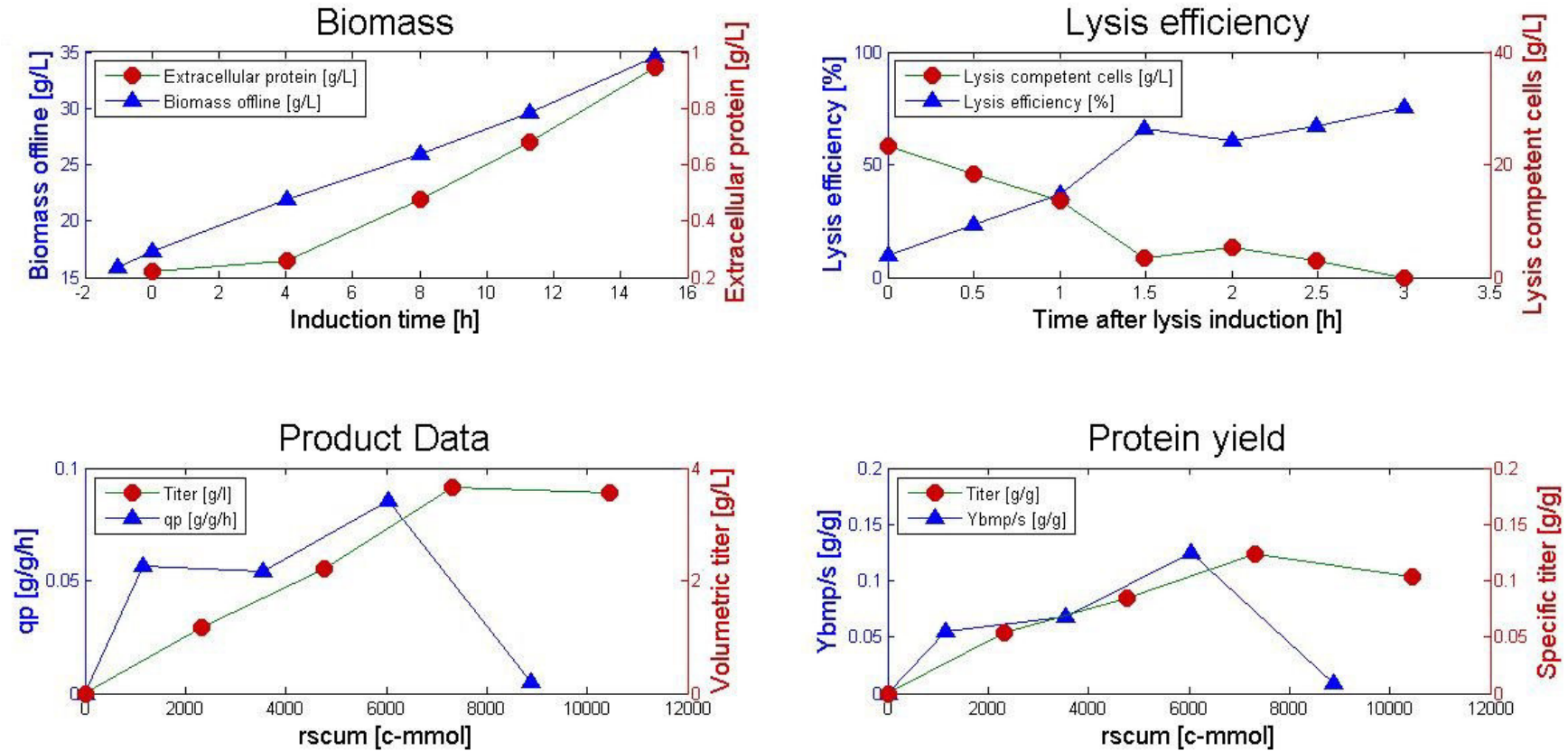


Figure 20: Offline data of fermentation 4. On the top left biomass concentration (blue) and extracellular protein concentration (red) are shown over induction time. The beginning of mixed feed was defined as starting point. Lysis efficiency (blue) and lysis competent cells (red) are plotted in the graph on the top right. The graphs on the bottom present specific rhBMP-2 production rate (blue) and the volumetric rhBMP-2 titer (red) on the left as well as the rhBMP-2 yield (blue) and the specific rhBMP-2 titer (red) on the right. On the abscissa in the two last mentioned graphs, there is cumulated substrate. With one exception offline data is only shown for induction phase. Lysis efficiency and lysis competent cells are plotted for lysis phase.

Extracellular protein

The lanes 5 and 10 in both gels in the gel in Figure 21 show the fermentation supernatant of fermentation 4, hence extracellular protein. Obviously there is one main protein in the extracellular fraction, which is found at approximately 50kDa. A typical protein in *E. coli* of this size is the homodimeric alkaline phosphatase, which is produced in the cytoplasm, and transferred to the periplasm of the bacteria, where the monomers form dimers through disulfide bond interaction. Its expression is associated with starvation – especially inorganic phosphate depletion – [117, 118] and stress [119]. The here conducted fermentations were run in a substrate limited way, which probably induced the stringent response due to carbon or amino acid depletion [120]. Hong et al. [118] proposed the stringent response to be a factor for the raise of alkaline phosphatase levels. Phosphate was only added into the batch medium; hence no further addition during fed-batch phase was done (see DeLisa et al. [108]). But the amount of phosphate added in form of potassium dihydrogen phosphate and di-ammonium hydrogen phosphate should not lead to phosphate depletions. Additionally at a sudden phosphate depletion a sharp increase of alkaline phosphatase would be expected, as the protein is not constitutively expressed, and its regulation is among other controlled by a phosphate-sensory protein [118]. As Figure 20 shows, extracellular protein concentration raised exponentially with increasing biomass concentrations, hence a sudden expression is barred. This growth associated formation of extracellular protein seems much more to be linked to permanent substrate limitation or a permanent stress due to recombinant protein production. As discussed above, the most common excreted protein found in the fermentation supernatant of these fermentations might be alkaline phosphatase. But to ensure this hypothesis it is necessary to investigate via enzyme activity tests or superiorly via circular dichroism, whether it is really the proposed protein. Furthermore it needs to be explained, how the protein is excreted into the medium. Before E-lysis induction, no common lysis was measurable and the amount of membrane deprived cells at the end of recombinant protein production phase (outlined in section 3.1.4) does not correlate with extracellular protein. It is possible that in some cells only the outer membrane is permeable, which leads to periplasmatic proteins release including the alkaline phosphatase.

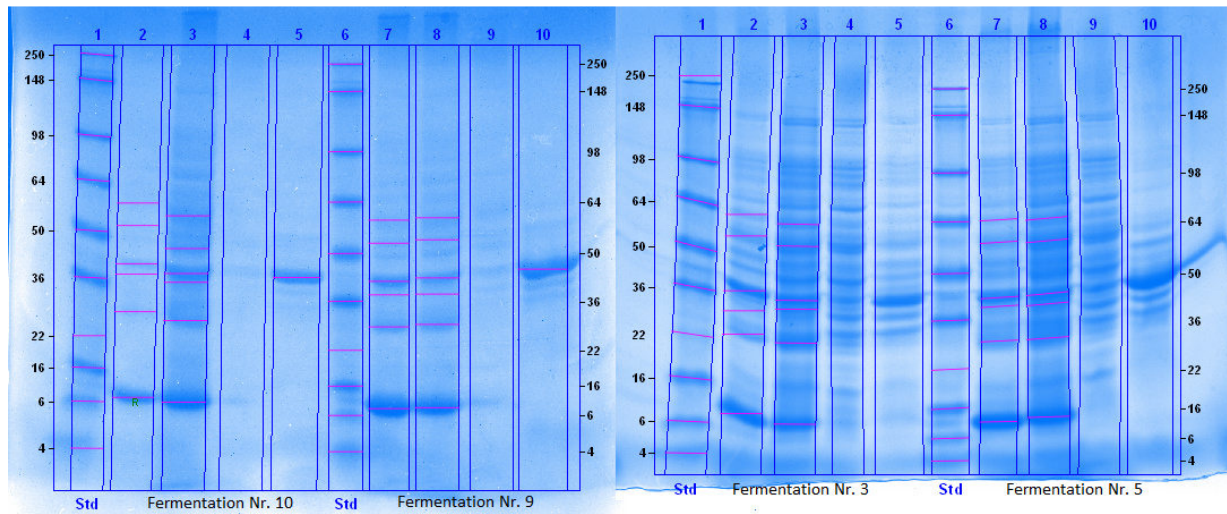


Figure 21: SDS-gel for fermentations 3, 5, 9, and 10. In lane 1 the SeaBlue Plus 2 Pre-Stained marker was run. For every fermentation four samples were run, whereas the samples for fermentation Nr. 10 are shown in the left gel in lanes 2-5 and Nr.9 on lanes 7-10. On the second gel fermentations 3 and 5 are presented. The samples for all fermentations have always the same order: Next to the ladder the homogenized cells from a sample taken before lysis induction, and next to it a sample of not-homogenized cells after three hours of lysis have been applied. The third lane contains the homogenisation supernatant of the last sample before lysis induction, and the lane on the right shows the fermentation supernatant of this sample.

Further samples shown in Figure 21 are discussed in section 3.3.

3.2.3 Processed data

Yield coefficients, rates and specific rates calculated using online and offline measured data are shown in Figure 23. The biomass yield coefficient ($Y_{x/s}$) as well as the other yield coefficients and rates staggered in the first 4 h of induction, but then reached a plateau. The $Y_{x/s}$ during induction phase was in average 0.37 ± 0.05 c-mol/c-mol in all eleven fermentations. It ranged from 0.31 c-mol/c-mol (FB Nr.9) to 0.48 c-mol/c-mol (FB Nr.7) in average (see Figure 22). The respiratory yield coefficients during induction phase for all fermentations had an average of 0.45 ± 0.05 mol/c-mol ($Y_{O_2/s}$), 0.49 ± 0.06 c-mol/c-mol ($Y_{CO_2/s}$) and 1.07 ± 0.04 c-mol/mol (RQ).

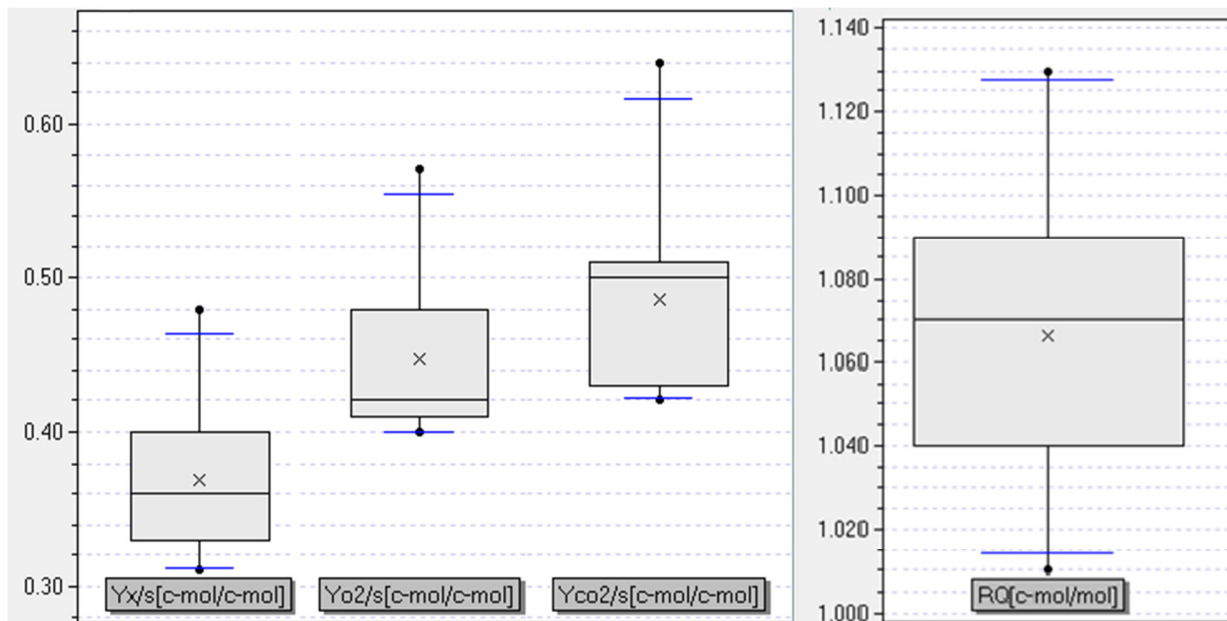


Figure 22: Mean s of $Y_{x/s}$, $Y_{o_2/s}$, $Y_{co_2/s}$ and RQ for multiple fermentations showed in boxplots. The box of the box-plot is confined by the first and the third quartile. The line in the box represents the median and the x the mean. The whiskers mark 1.5 times o the interquartile distance.

All yield coefficients, rates and specific rates seem to need some time for adaption at the beginning of induction phase. After the first four hours the trends are more constant than the hours before. Abrupt changes in biomass yield coefficients and growth rate shown in Figure 23 are due to biomass DCW values. Samples were taken in certain intervals. In-between two samples a continuous trend occurred and changed abruptly dependent on the value of the next sample.

The time to adapt the cells might be due to the change in the feeding profile. Before the start of L-arabinose feed, L-arabinose pulse is given, where the cells receive more substrate than they are able to aerobically metabolize. When the L-arabinose is consumed, a substrate limited phase starts. Substrate limitation is associated with the stringent response [23]. Studies showed that at the start of fed-batch after a batch phase factors associated with the stringent response increased. After some time of adaption on the these factors goes back to its initial level, whereas ppGpp rests elevated [121]. Furthermore, an adaption of glucose affinity was reported [122].

Three hours after induction specific substrate uptake rates are slightly increasing until the end of recombinant protein production. This might be due to the overestimation of the biomass

concentration with progressing induction time. An overestimation of biomass induces a too high feed-rate compared to the set-point of total q_s .

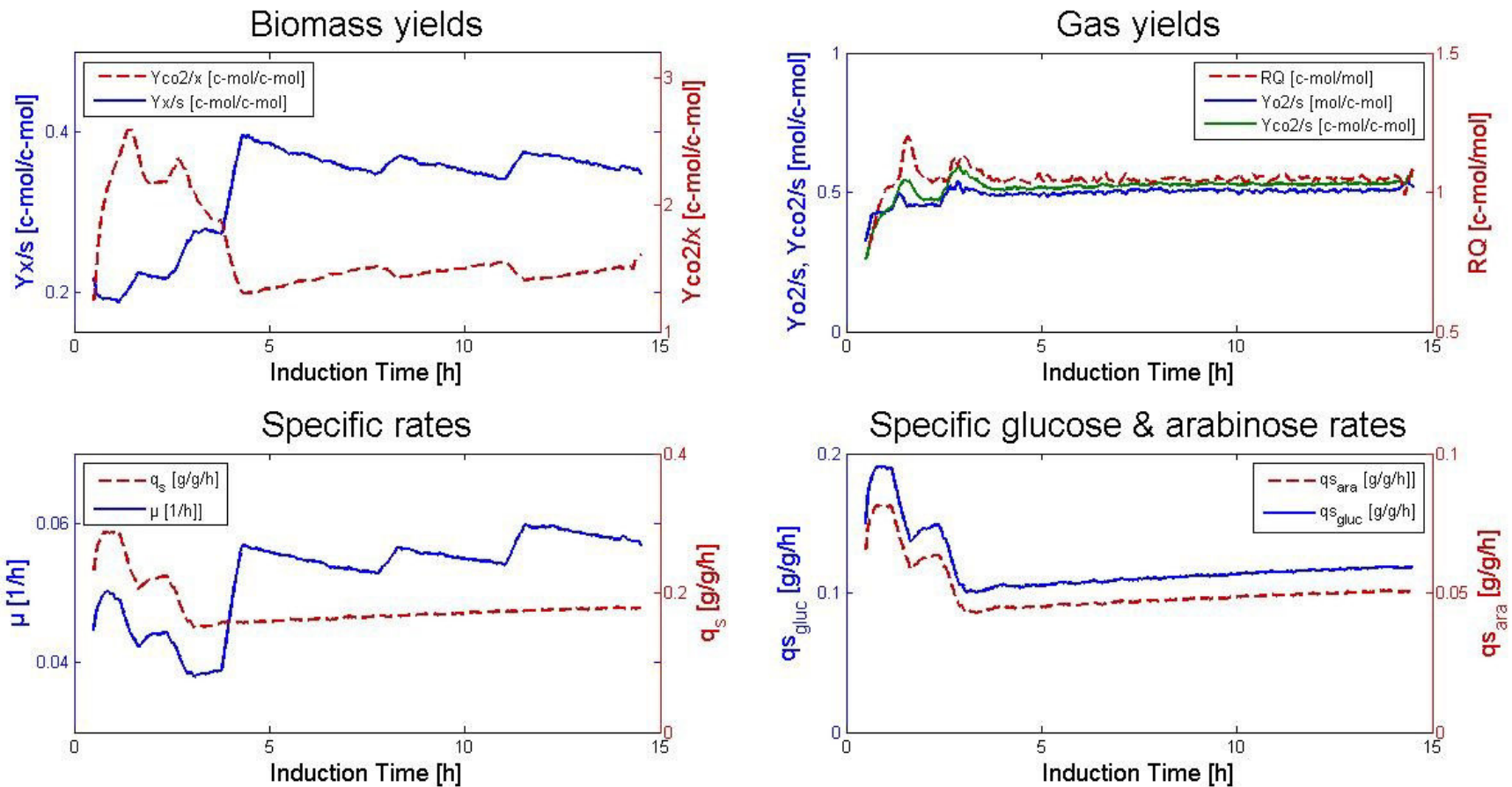


Figure 23: Processed data of fermentation 4. Biomass yield (blue) and $Y_{CO_2/s}$ (red) are plotted in the graph on the top left. Next to it the gas yields are shown ($Y_{O_2/s}$ in blue, $Y_{CO_2/s}$ in green and the respiratory quotient in red). On the bottom the total specific substrate uptake rate (red) and the specific growth rate (blue) during recombinant protein production are found in the left, whereas specific D-glucose uptake rate (blue) and specific L-arabinose uptake rate (red) are on the bottom right. Processed data is only shown for the induction phase.

3.2.4 Information matrix

Online, offline and processed data were reduced to one single value per variable and fermentation. Therefore either the average over time or a defined value (e.g. last sample of induction phase) was taken. A section of the information matrix is shown in Table 5. The whole matrix is found in the appendix in Table 7 to Table 10.

Table 5: Section of the information matrix

Fermentation Nr.	FB.Nr.	qs gluc SP [g/g/h]	qs ara SP [g/g/l]	qs tot SP [g/g/l]	xAra SP [%]	μ [1/h]	Y _{x/s} [c-mol/c-mol]	RQ [c-mol/mol]
26	4	0.125	0.050	0.175	28.57	0.05	0.33	1.05
27	2	0.200	0.125	0.325	38.46	0.12	0.32	1.03
29	7	0.125	0.125	0.250	50.00	0.10	0.48	1.08
30	8	0.125	0.200	0.325	61.54	0.12	0.37	1.07
31	1	0.200	0.050	0.250	20.00	0.08	0.34	1.09
32	10	0.050	0.125	0.175	71.43	0.07	0.36	1.12
33	9	0.075	0.075	0.150	50.00	0.05	0.31	1.04
34	3	0.200	0.200	0.400	50.00	0.16	0.40	1.01
35	5	0.125	0.125	0.250	50.00	0.09	0.40	1.13
36	6	0.125	0.125	0.250	50.00	0.10	0.40	1.04
38	1	0.200	0.050	0.250	20.00	0.10	0.35	1.07
39	12	0.290	0.110	0.400	27.50	0.15	0.43	1.01

3.2.5 Multivariate data analysis

Principal component analysis

The information matrix consisting of 40 variables was load into Matlab (MathWorks, Natick, Massachusetts, USA), DataLab (Epina GmbH, Pressbaum, Austria) and Modde (Umetrics, Umea, Sweden) for further statistical investigation.

As first step a PCA was done to investigate in linkages between the variables. In Figure 24 a bi-plot shows the original variables in the space of principal components 1 to 2. The PCA was calculated excluding the data of fermentation 7, as the culture in this run was induced at 25g/l instead of 15g/l biomass concentration. The first principal component (PC) explained variance of 49.43%, the second and the third one 20.92% and 13.25%. The fourth PC only explained 5.59% and also the third PC with 13.25% rather seemed to explain noise than variance induced by process parameter changes. Together the first two PCs explained 70.34% of total variance.

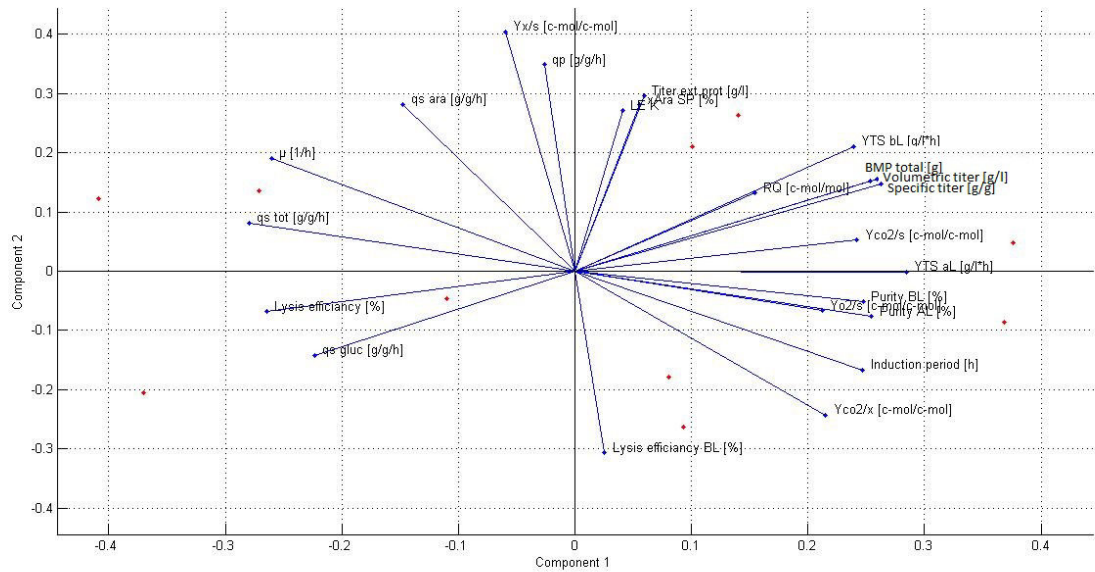


Figure 24: 3-dimensional bi-plot showing the variables on principal components 1-2

The variables with the highest positive loadings on PC1 were qs_{gluc} , qs_{tot} , the specific growth rate and lysis efficiency. Negative loadings had induction period, multiple variables concerning rhBMP-2 titer, purity and $Y_{CO_2/s}$ (see Figure 25). Highest loadings on PC2 were all positive. The variables were qs_{ara} , X_{ara} , $Y_{x/s}$, qp and titer of extracellular protein.

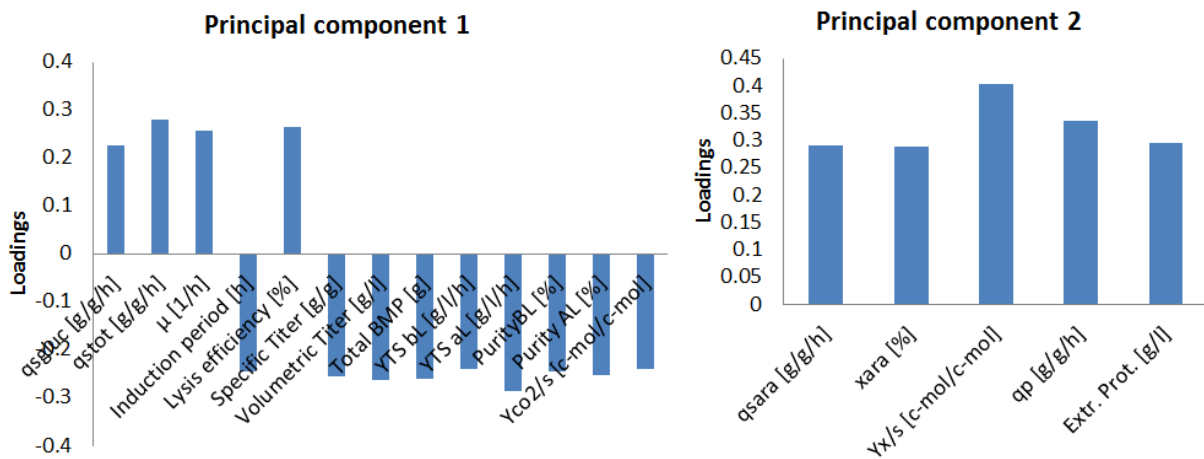


Figure 25: Variables with the highest loadings in principal component 1 (left) and 2 (right)

Hence the first PC represents metabolic load and cell stress represented by product titer and the ability to lyse. On the other hand the second PC could be interpreted as representing the pBAD system, by the inducer arabinose (“trigger of recombinant protein production”) and the

rhBMP-2 production rate. In this way the two PCs represent the CPPs of the here conducted DoE. The specific D-glucose uptake rate highly loads on the first PC, and the specific L-arabinose uptake rate on the second PC. But as the loadings on PC1 and PC2 show, the original CPPs could be transmitted to total specific substrate uptake rate and L-arabinose concentration instead (see Figure 26). With these two parameters the design space could be described in the same manner as it was done with specific L-arabinose and D-glucose uptake rate.

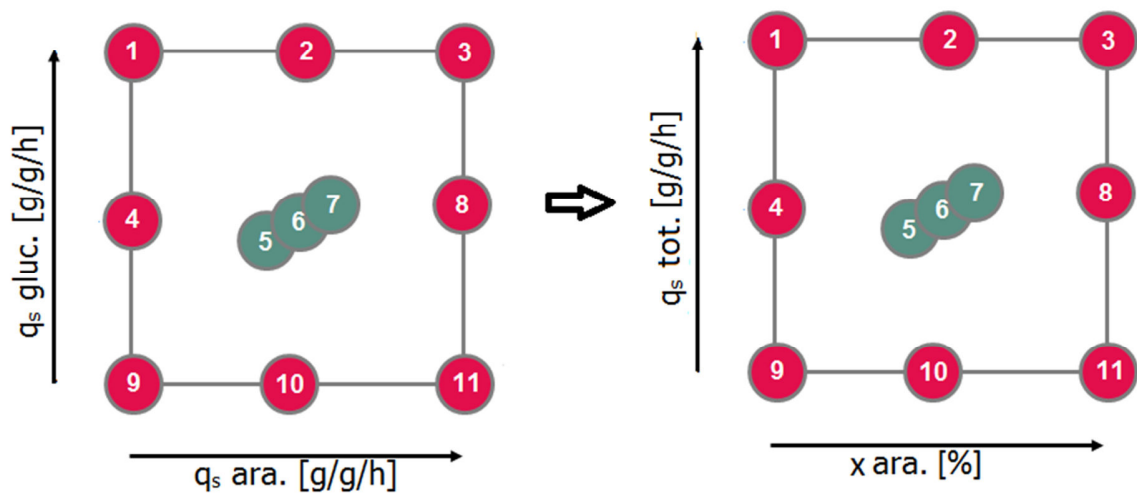


Figure 26: Transfer from the original DoE using specific L-arabinose and D-glucose uptake rate as CPPs (left) to a changed DoE with total specific substrate uptake rate and concentration of L-arabinose in the feed as CPPs instead

Due to the loading plots, high total specific substrate uptake rate which is collinear to growth rate and induction period is distinct to achieve high lysis efficiency. Opposed to this, lower specific substrate uptake rate seems favorable for high rhBMP-2 titers. Concerning the second PC, a high percentage of L-arabinose in the feed is associated with high specific rhBMP-2 production rates. This goes in line with the triggering of recombinant protein production by the amount of L-arabinose fed. Furthermore, a low total specific substrate uptake rate seems to be associated with higher rhBMP-2 purity and higher $Y_{CO_2/s}$. This could be caused by the appearance of metabolic burden, due to raised recombinant protein titers. The cell needs more energy proportional to lower recombinant protein production, and this means a higher CO_2 emission [32]. But also maintenance metabolism, were due to lacking resources the C-source is mainly used for energy supply without biomass formation might have occurred [123].

Due to PC2 high L-arabinose levels go in line with high biomass yield coefficients and extracellular protein content. For the second mentioned, the above discussed theory about the raise of alkaline phosphatase caused by the adaption of *E. coli* to stressing environmental conditions could be the reason.

The high biomass yield coefficient at high L-arabinose rates is against the findings discussed in section 1.7.3. It was hypothesized that $Y_{x/s}$ decreases with raising specific L-arabinose uptake rates, because L-arabinose rate is assumed to be proportional to the specific rhBMP-2 production rate, which was confirmed by the PCA. High recombinant protein formation rates go in line with higher energy demands, which is associated to increasing $Y_{CO_2/s}$ and hence decreasing $Y_{x/s}$ [29, 32]. Furthermore, D-glucose as the preferred carbon source should be better degradable for the bacterium than L-arabinose. And the higher the L-arabinose concentration in the feed, the lower is the D-glucose concentrations.

The calculation of $Y_{x/s}$ bases on biomass DCW measurements. Inclusion bodies are very dense particles which may raise the average density of cells. Hence the mass of the cells increased, resulting in higher DCW and further higher biomass yield coefficients.

Multilinear regression

For further investigation of the connection between variables showed by the PCA, multilinear regressions were calculated in Modde. The process parameters total specific substrate uptake rate and the specific rates for L-arabinose and D-glucose are according to the experimental design not collinear (orthogonal design), as also shown in the PCA plot. Hence, they can be together used as independent variables for regression calculations, without impeding the interpretability of the model coefficients. Total q_s is highly correlating with induction period and growth rate, and therefore were more or less exchangeable as independent variables. The same is true of the specific L-arabinose uptake rate and L-arabinose concentration in the feed.

MLR results for lysis efficiency, q_p , titer and purity will be outlined in the following sections. Neither on extracellular protein titer nor on the biomass yield significant influences were found. As opposed to this a linear regression was fitted on $Y_{CO_2/s}$. The negative impact of $q_{s_{tot}}$ on this yield was confirmed by a model having a R^2 of 0.816 and a Q^2 of 0.714 (see Figure 27). For the

fitting of the model fermentation 10 had to be excluded. This could be reasoned by the fact that during the induction phase of this fermentation, the exponential feed was interrupted by 90 min of constant feed. Even though data of this constant feeding phase was not included in the calculation of average $Y_{CO_2/s}$ over induction phase, it might have an influence. The link between total specific uptake rate and $Y_{CO_2/s}$ was discussed in the previous section.

For $Y_{O_2/s}$ a MLR was calculated. A negative effect of $q_{s,gluc}$ was found and additionally a quadratic effect of this parameter. The R^2 was 0.811, and the Q^2 0.627. Furthermore, the model had a validity of 0.517 and a reproducibility of 0.916. The negative connotation between specific D-glucose uptake rate and $Y_{O_2/s}$ means that the higher $q_{s,gluc}$ was the less oxygen was consumed for one mole of substrate. Hence the more D-glucose there is in the feed, the more efficient is the utilization of oxygen.

The quadratic effect suggests a $q_{s,gluc}$ point, at which the most oxygen is needed to metabolize the substrate. Lower as well as higher specific D-glucose uptake rates need lower oxygen levels per mole substrate.

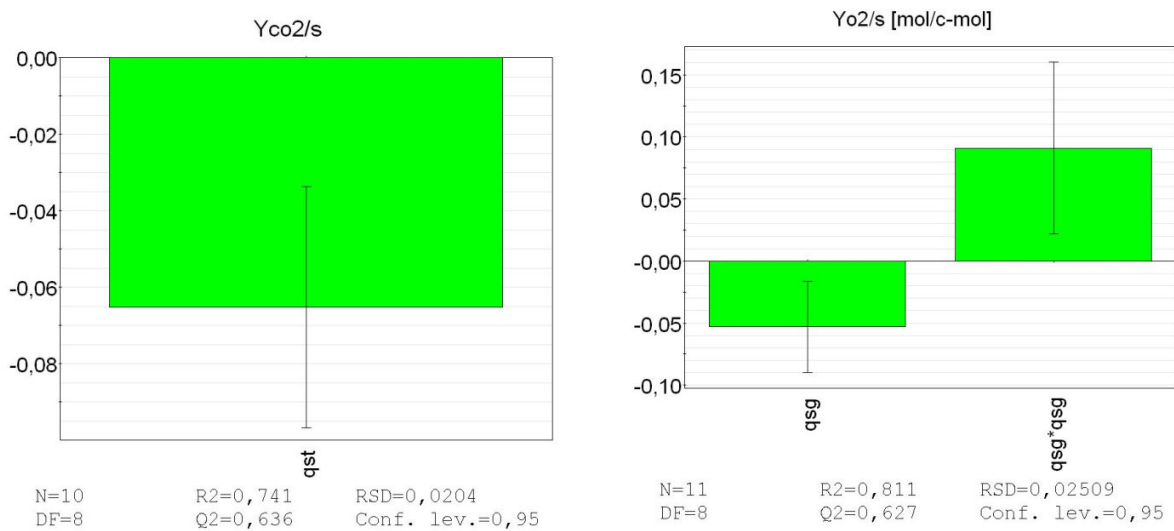


Figure 27: Regression model for $Y_{CO_2/s}$ (left) and for $Y_{O_2/s}$ (right)

3.1 E-lysis of *E. coli*

3.1.1 Flow cytometry: The method

As outlined in section 2.2.6, the method to measure BGs via flow cytometry had been adapted from Langemann et al. [7]. When comparing the measurements of Langemann et al. [7] with the measurements in this study, it is apparent that in this study the BGs in the space of FSC and FL1 are at the place where Langemann et al. [7] found dead cells. Their BGs had a smaller value in FSC.

The method had been further developed and the measurement of dead cells investigated. (Dead cells are defined as dead cells which are not E-lysed). These cell population shift up higher in FL1, than it had been shown in Langemann et al. [7]. Hence BGs and dead cells can be more clearly distinguished. The difference in the FL1 signal between these two cell populations might be due to the effect of DiBAC4(3). The fluorescence of this stain was shown to be dependent on its interaction with other molecules, like proteins in the cytoplasm [124]. As BGs are expected to be empty, the fluorescence might be not as strong as in dead cells, which are still intact.

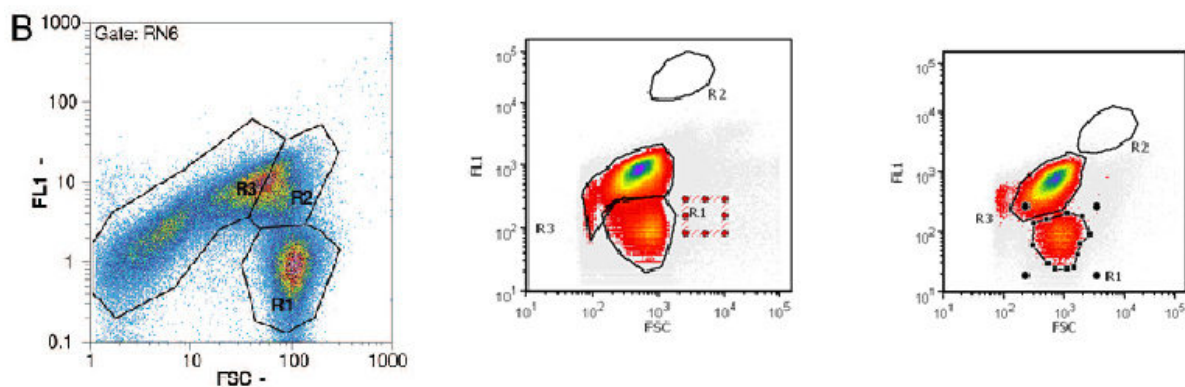


Figure 28: Comparison of the flow cytometry method of Langemann et al. (2010) and as it was conducted in this study

The shift to the left in the FSC, which is associated with the discharge of the cell content, hence the emptiness of the cells, might arise from inclusion body production in the here conducted study. It is possible, that the inclusion bodies are too big to pass the hole in the cell wall. Hence the cell is not totally empty. Furthermore, the cell size of *E. coli* is dependent on the growth rate [125]. Hence the cells in a batch culture are predicted to be bigger than in substrate limited

cultures, as cells during induction phase in this study were. Therefore, the difference in size between E-lysed cells and cells after induction phase in this study might be not as distinct as in Langemann et al. [7].

3.1.2 Lysis efficiency

Lysis efficiency is defined as the percentage of bacterial ghosts in the whole cell population. As outlined in section 2.2.6 it was investigated via flow cytometry. Lysis efficiencies varied between 31% in fermentation 10 and 99% in fermentation 2. The trend of lysis for these two fermentations and for lysis of fed-batch 1 is shown in Figure 29.

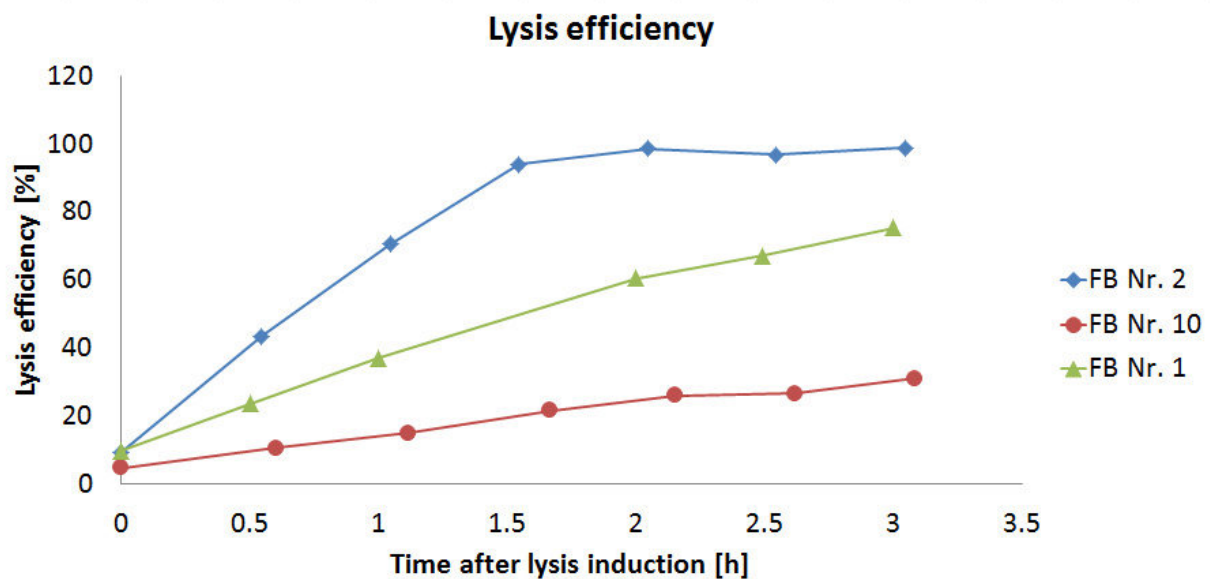


Figure 29: The trend of lysis efficiency over induction phase for fermentations number 1, 2 and 10

Lysis efficiency at the end of lysis phase was investigated. It was shown to have a high loading in the PCA described in section 3.2.5. Due to this PCA the percentage of L-arabinose in the feed should correlate negatively with lysis efficiency, whereas the total specific substrate uptake rate should do positively. Figure 30 shows these correlations.

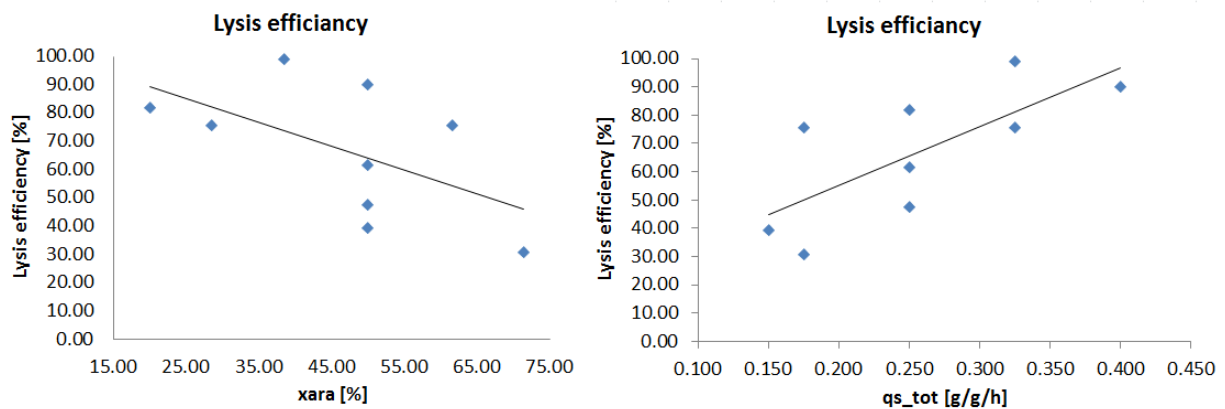


Figure 30: Correlation of lysis efficiency with L-arabinose concentration in the feed (left) and total specific substrate uptake rate

An MLR with L-arabinose concentration in the feed and $q_{s_{tot}}$ as independent as well as lysis efficiency as dependent variable was conducted. Figure 31 shows the positive influence of $q_{s_{tot}}$ on lysis efficiency and a negative for L-arabinose concentration in the feed. 89.5% of the variance is explained by this model. Q^2 has a value of 0.766.

The MLR as well as the correlations showed in Figure 30 were done excluding fermentation 1a and 7. The first of these two fermentations was oxygen depleted at the onset of lysis phase, which induces cell stress and changed physiology of the cells by a shift of totally aerobic to partially anaerobic metabolism. This is supposed to have a negative influence on lysis efficiency, because the ability to E-lyse is dependent on the physiological state of the cells [6, 7]. In fermentation 7 the culture was induced at 25g/l biomass concentration instead of 15g/l. Hence induction phase was shorter than it was in the other processes.

The model predicts a lysis efficiency of at least 90% when $q_{s_{tot}}$ is higher than 0.25g/g/h and the L-arabinose concentration in the feed maximal 55%. For lysis efficiencies of more than 95%, minimal $q_{s_{tot}}$ are 0.3g/g/h with 40% L-arabinose in the feed at the highest.

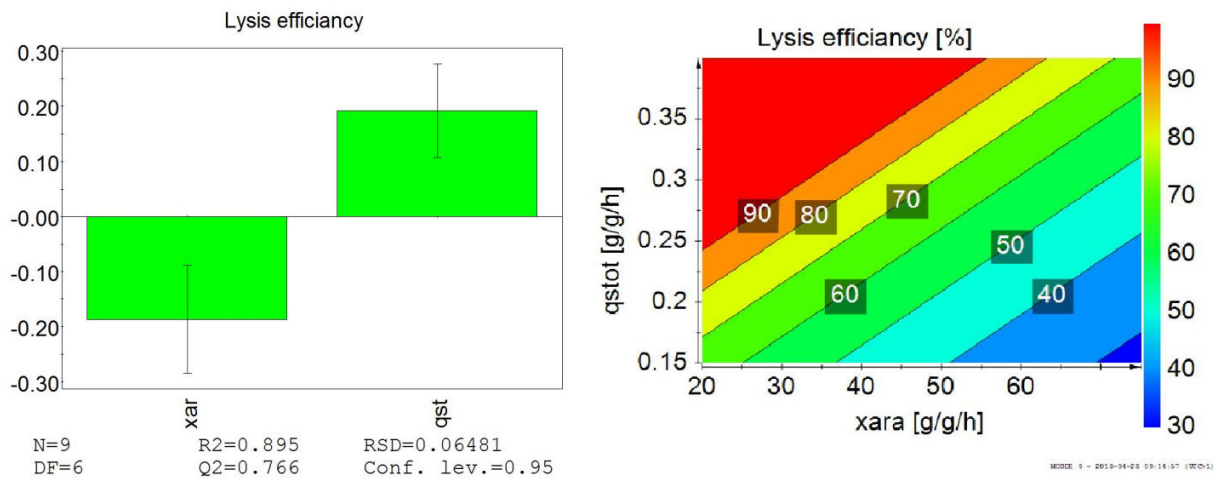


Figure 31: The coefficients of the MLR conducted for lysis efficiency (left). On the right a prediction for lysis efficiency dependent on the concentration of L-arabinose in the feed and q_{stot} is showed.

Total specific substrate uptake rate as factor is equivalent to induction period and growth rate, as these three variables are co-linear. This can be explained by the fact that the higher the total specific substrate uptake rate is, the higher is the growth rate. Furthermore, the fermentations were induced at 15g/l biomass concentration and induction phase was stopped at 35g/l. Hence the induction period was the shorter, the higher the total specific substrate uptake rate and the growth rate were. Figure 32 shows the negative influence of induction time on lysis efficiency, and the positive impact of the growth rate.

The colinearity of total specific substrate uptake and induction period was the reason to not consider fermentation 7 in the MLR calculation. In this fermentation the induction phase was shorter, as the culture had been induced at too high biomass concentrations. Its lysis efficiency of 58% fitted to the values for fermentations 5 and 6, which had the same $q_{s_{ara}}$ and $q_{s_{gluc}}$ set-points, and which reached lysis efficiencies of 62 and 47%. Hence the total specific substrate uptake rate might be more distinct than the induction period. But on the other hand it could also be that the whole fed-batch phase (induced and un-induced) could have an influence, which would mean that in this concern there was no difference between the three fermentations. The lysis efficiency of fermentation 6 was low compared to the other two, but was still in-between the variation range of the model. To effectively distinguish the effects of total specific substrate uptake rate and induction period, the culture needs to be treated in exactly the same way before being induced.

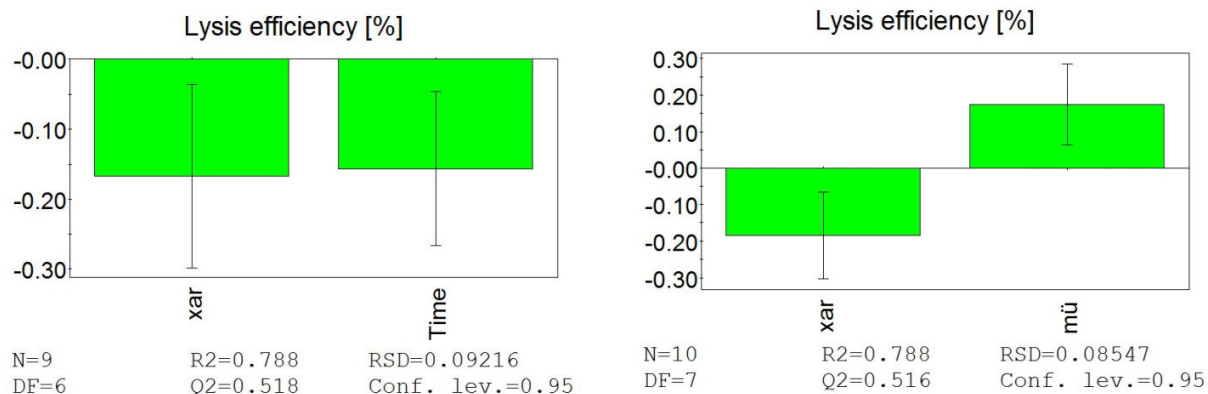


Figure 32: L-arabinose concentration in feed and induction time (left) as well as L-arabinose concentration in feed and the growth rate (right) as coefficients in an MLR to explain lysis efficiency

The PCA shows a negative connection between lysis efficiency and rhBMP-2 titer. As so far investigated, high L-arabinose concentration in the feed has a negative impact on lysis efficiency. This was explained by metabolic burden due to higher recombinant protein production with raising L-arabinose concentrations in the feed. Hence the titer of rhBMP-2 is expected to have a negative impact on the ability to E-lyse.

This expectation was confirmed by a MLR using the specific titer as independent, and the lysis efficiency as dependent variable (Figure 33). 61.8% of the variance in lysis efficiency can be explained by differences in the specific titer. The Q^2 was 0.446. The model validity was at 0.81 and the reproducibility at 0.78.

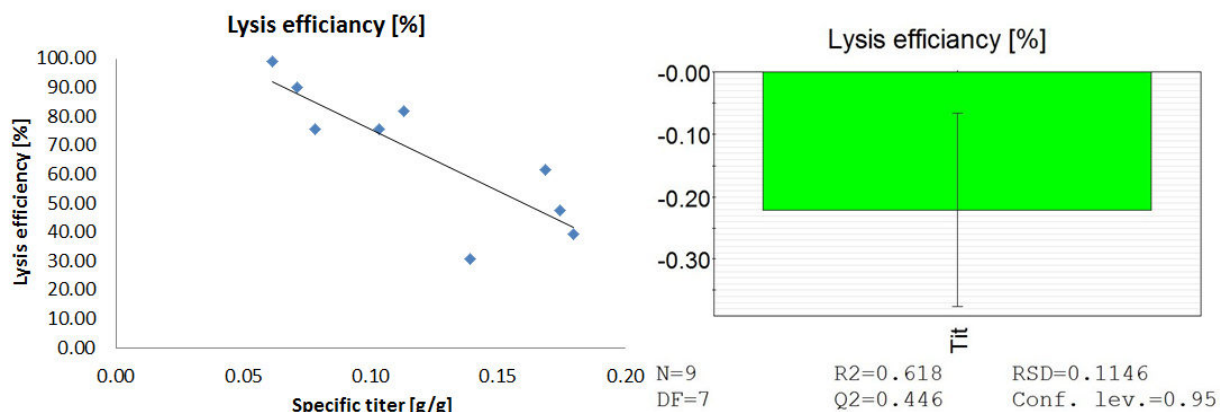


Figure 33: The correlation between specific titer and lysis efficiency (left) as well as the MLR coefficient plot for lysis efficiency

The findings concerning lysis efficiency at the end of fermentation can be explained by a high metabolic load due to recombinant protein production [116]. As the assumption was made,

that the amount of L-arabinose in the feed triggers the rate of induction, a higher L-arabinose concentration in the feed is associated with higher recombinant protein expression rate and hence higher metabolic load. The stress on the cells may alter their physiology in a way to decrease the ability to E-lyse. This goes in line with the negative impact of longer induction periods on lysis efficiency, as longer induction time means a longer period where the cell is exposed to stress. Furthermore, cell segregation during energy limited growth [126] and recombinant protein production [127] could have an impact on lysis efficiency. Thus at too low growth rates and specific substrate uptake rates, the cells are energy limited which is coherent with cell segregation into viable but non-dividing cells [126]. Cells, which are in a damaged physiological state that they are no more able to divide, may neither be able to lyse.

3.1.3 Lysis kinetics

E-lysis competent cells have been described in section 3.2.1. They were calculated depending on the percentage of cells in the culture lysed.

The model to explain lysis kinetics bases on this cell population. Compared to the lysis efficiency itself, it was advantageous to use the population of E-lysis competent cells to fit a model, because there - beside of lysis efficiency - the biomass concentration at lysis onset, as well as the segmentation of the culture into cells of varying physiological states was considered. The cell populations, at the end of induction phase, were proposed to consist of culturable cells (which are supposed to be E-lysis competent) and viable non-culturable cells. The second was expected to emerge because of cell segregation due to substrate limitation [126] and recombinant protein production [127]. These two sub-populations with their behavior during lysis phase are sketched in Figure 34.

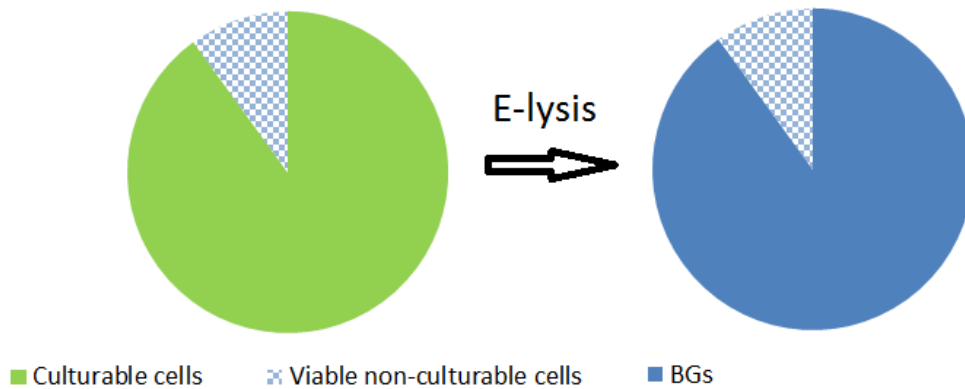


Figure 34: Culturable and viable non-culturable cells before and after E-lysis

In addition to Figure 29 showing the lysis efficiency over lysis phase for fermentations 1, 2 and 10, Figure 35 presents the trend of E-lysis competent cells for the same fermentations in the same time period. The number of E-lysis competent cells decreases with ongoing lysis phase, until it reaches a value of zero at the end of lysis.

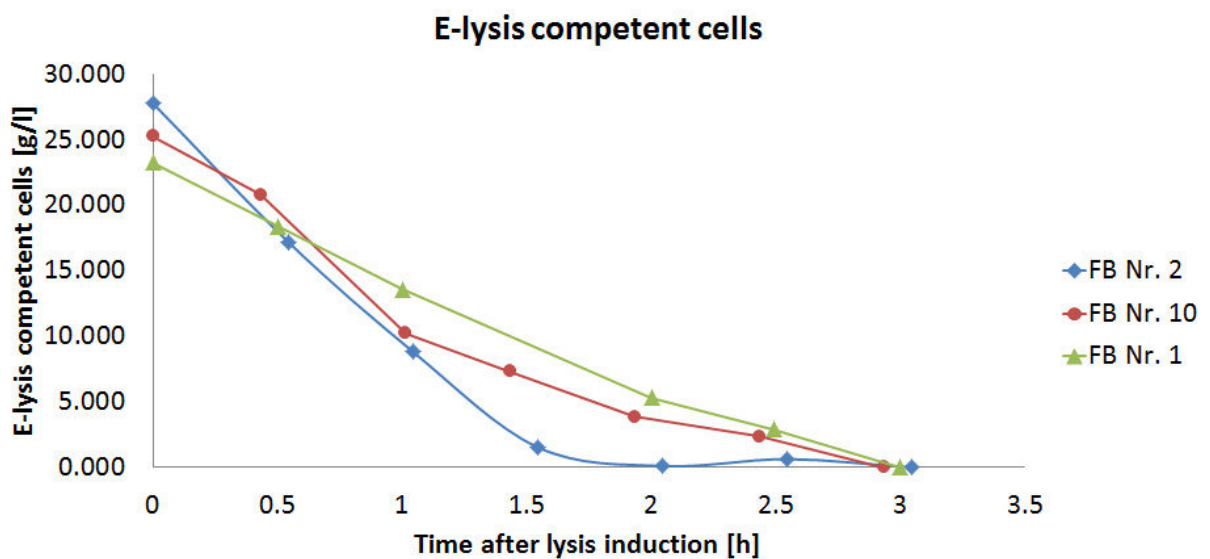


Figure 35: The trend of E-lysis competent cells of fermentations 1, 2 and 10 during induction phase

A negative exponential curve showed to appropriately fit the trend of E-lysis competent cells over time. Hence the amount of E-lysis competent cells (N) over time (t_L) can be described as Equation 11 shows with k being the reaction constant.

$$\frac{dN}{dt_L} = k * t_L$$

Equation 11: First order kinetics describing the trend of E-lysis competent cells over lysis phase

The exponential curve was fitted using a curve fitting tool in Matlab. As example of a fitted curve, the trend of E-lysis competent cells for fermentation 2 is shown in Figure 36.

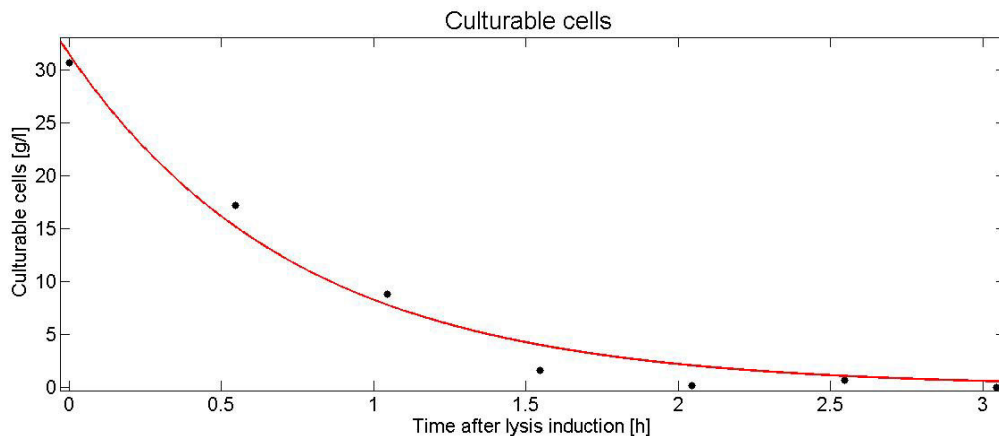


Figure 36: E-lysis competent cells during lysis. A curve fitting was done to explain the kinetics of lysis.

The curves for all fermentations were fitted to have the highest possible accordance with the measured points. The reaction constants varied between -0.97 and -0.47, and additionally there was one outlier with a k of -1.34 (see Figure 37). Interestingly, this outlier belonged to fermentation 2, which had the highest lysis efficiency of all fermentations - 99%.

The mean reaction coefficient was -0.76 ± 0.29 . This value can be used to predict the time of lysis needed to reach an aspired lysis efficiency, for varying biomass concentrations. This is important for future fermentations, especially in a probable manufacturing process, to plan lysis phase at the end of recombinant protein formation.

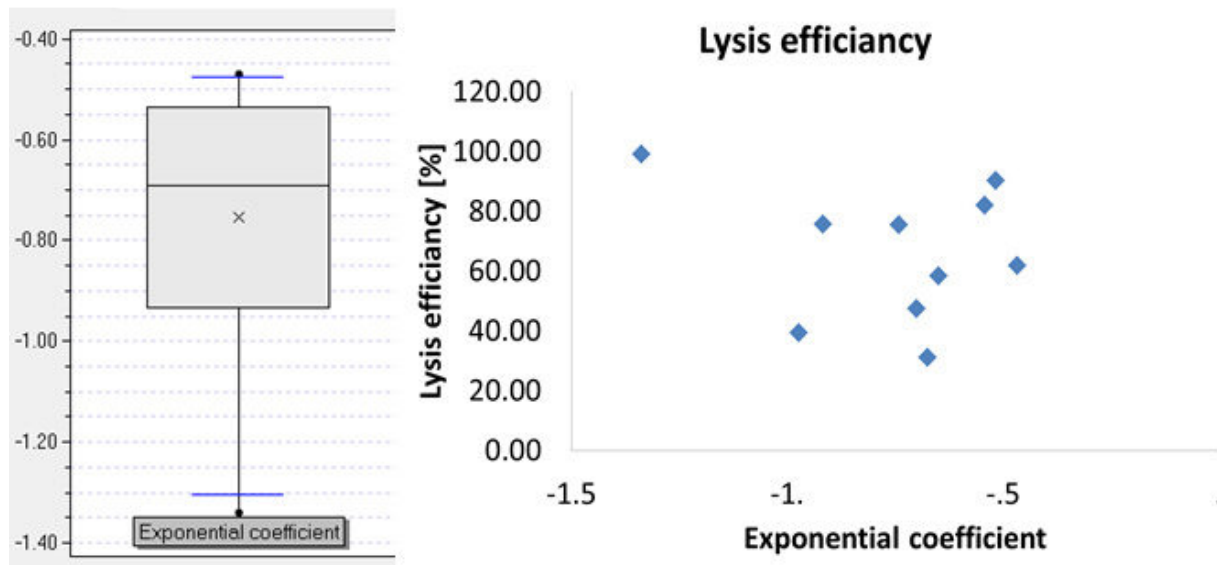


Figure 37: Distribution of reaction constant k over multiple fermentations (left) and the correlation between this coefficient and lysis efficiency

There was neither a correlation between lysis efficiency and lysis kinetics found (see Figure 37) nor another parameter which influences k significantly. Therefore it is not possible to predict k depending on process parameters. But an average k considering its error can be applied to plan lysis phases and therefore simplify fermentation processes. Hence under physiologically satisfying conditions for protein E mediated lyses (outlined in section 1.4) this k value can be used to estimate the time to full lysis.

3.1.4 Subpopulations during mixed feed phase 2

Membrane deprived cells, which are defined as cells with higher fluorescence after DiBAC₄(3) staining than living cells. In all fermentations these cells – being detected as BGs in the flow cytometer - were already detected before the temperature shift to 42°C. But before lysis induction no BGs were expected, except in case the expression system is leaky. It was hypothesized whether these cells were viable but non-culturable. Sorenson et al. (2004) distinguished living from viable non-culturable via flow cytometry using DiBAC₄(3). The utilization of this stain was justified by the fact that the membrane potential is linked to energy, as it is driven by the passage of charged molecules through the membrane. In viable non-

culturable cells the membrane potential is expected to be collapsed [14]. In this study living cells have been distinguished from BGs via DiBAC₄(3) staining and the forward scatter signal. Figure 38 shows some samples taken before lysis induction. A part of the sample (marked by an arrow) is more intense in FL1 meaning, that the membrane potential is lower. The signal in the FCS did not change. These cells were wrongly marked as BGs, but also could be interpreted as viable non-culturable cells.

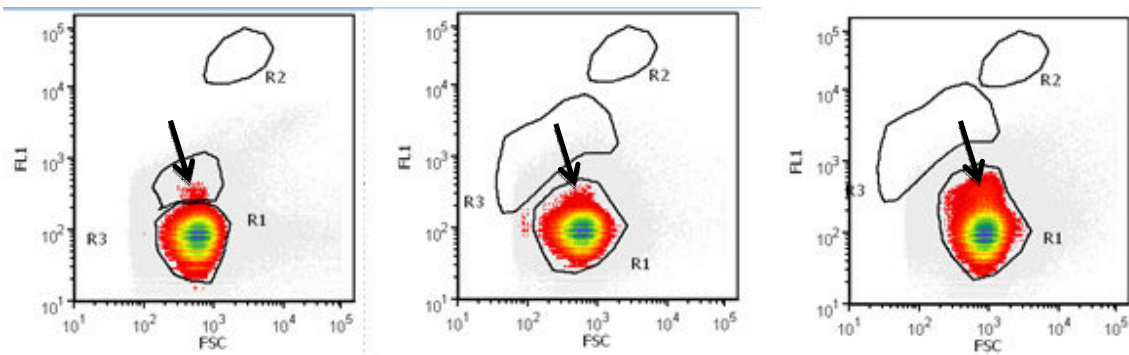


Figure 38: Flow cytometer measurements before lysis induction. The second picture was a sample in the same fermentation as the third picture, but taken 30min earlier

Due to the PCA in section 3.2.5 the number of membrane potential deprived cells before lysis induction has a negative dependence on $q_{s_{ara}}$ as well as of the concentration of L-arabinose in the feed. This was confirmed with an MLR, where the L-arabinose concentration in the feed explained 72% of variance in lysis efficiency before lysis induction (see Figure 39). The validity of the model was 0.84 and the reproducibility 0.68.

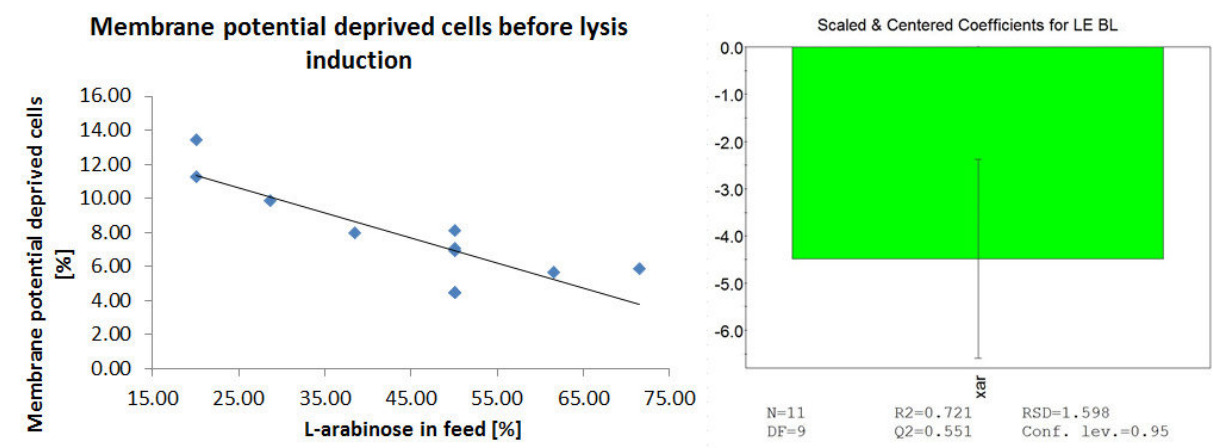


Figure 39: The correlation between L-arabinose in the feed and the membrane potential deprived cells before induction (left) as well as the MLR model (right)

The correlation between L-arabinose in the feed and membrane potential deprived cells before lysis induction had been expected in the opposite direction. Hence an increasing concentration of L-arabinose in the feed goes in line with higher metabolic load, which was expected to have a higher amount of viable non-culturable cells in the population. But as the correlation is negative, L-arabinose seems to have a positive influence to not disturb the membrane potential.

Furthermore, viable non-culturable cells would be expected to not be able to E-lyse. Figure 40 shows the trend on the flow cytometer for the lysis phase of fermentation 2. The sample on the left has been taken before lysis induction. In the picture the whole cell population is marked as living cells. But when a closer look is taken, the top of R1 attracts attention. A part of the population has shifted up, and thereby should rather be marked as BG or better viable non-culturable. But when the sample after three hours lysis is regarded, this “viable non-culturable” population is disappeared. The correlation of the observed pre-lysis subpopulation with $q_{s_{ara}}$ is still inconclusive; probably there is an unidentified underlying variable. Further studies will be necessary to elucidate this observation.

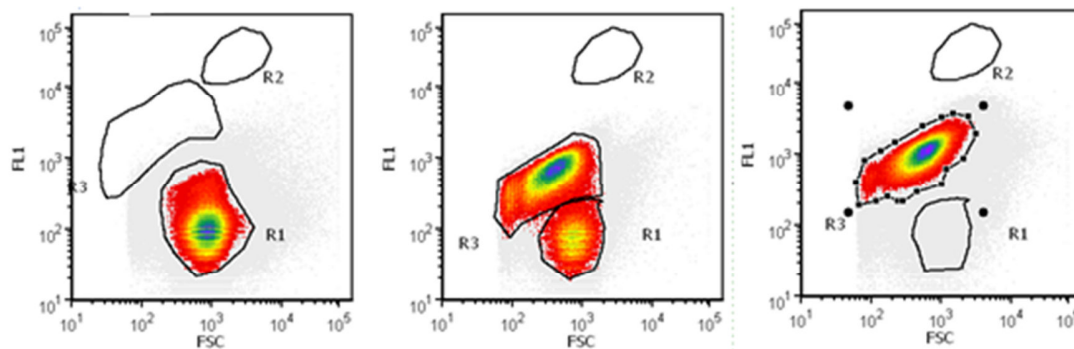


Figure 40: Flow cytometer measurements of FB Nr. 02. The picture on the left shows the culture before lysis induction, the second picture 60 min later and the third measurement was done 3 h after the temperature shift.

3.2 *rhBMP-2 formation*

In all processes the specific *rhBMP-2* production rate rose at the beginning to then reach either a plateau or to sink after a certain point. The three fermentations where the production rate started to decrease were the low-low (FB Nr. 9) and the high-high (FB Nr.3), as well as FB Nr. 2

with a $q_{s_{tot}}$ of 0.325 and the highest lysis efficiency. In all other fermentations it stayed at a constant level.

In Figure 41 trends of specific rhBMP-2 production rates are shown. On the left differences in trend of q_p triggered by total q_s are presented. On the right side the influence of specific L-arabinose uptake rate is lined out. The PCA in section 2.3 proposes a connection between these specific L-arabinose uptake rate and specific rhBMP-2 production rate, but via MLR no significant interaction was found. Neither another process variable showed to significantly influencing q_p . Due to the data achieved in this study, which is partially shown in Figure 41, a quadratic effect of $q_{s_{ara}}$ on q_p is hypothesized, as also observed elsewhere in this system [8]. This would mean that at average values of $q_{s_{ara}}$, the level reached by the specific rhBMP-2 production rate is the highest. To statistically confirm the model, further investigations would be necessary.

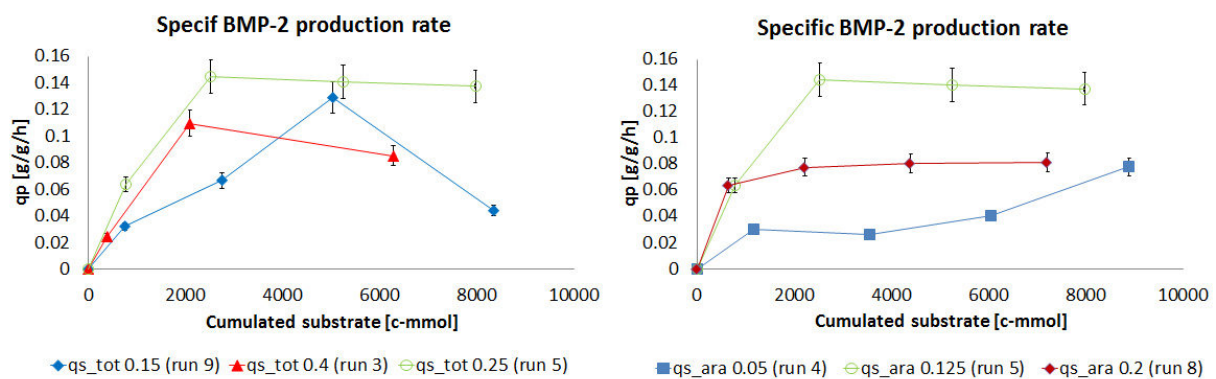


Figure 41: q_p trends over a range of different $q_{s_{tot}}$ (left) and over differing $q_{s_{ara}}$ (right)

When the specific titer instead is investigated, the graphs in Figure 42 show how total specific substrate uptake rate and specific L-arabinose uptake rate influence this parameter. For the first mentioned, a high specific uptake rate seems to have a negative impact on product formation. For the specific L-arabinose uptake rate an average value has the highest rhBMP-2 titer.

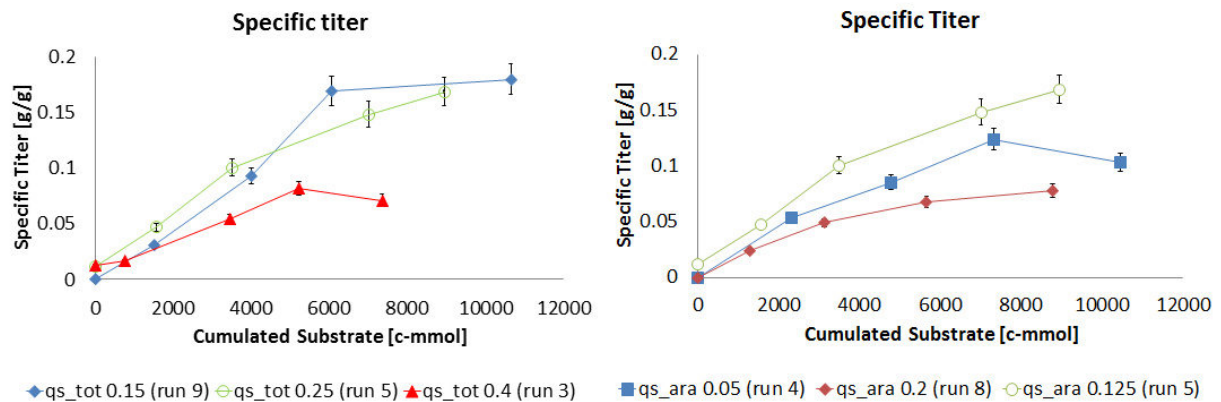


Figure 42: Specific titer trends for varying total specific substrate uptake rates (right) and for $q_{s_{ara}}$ (right)

These observations were investigated in a MLR, and the total specific substrate uptake rate solely turned out to influence the specific titer (see Figure 43). The model predicts that the specific titer is the higher the lower total q_s D-glucose is. The model explains 50% of the variance in specific titer. The prediction is with a Q^2 of 0.292 rather low. Fermentation 7 had been excluded due to its short induction period.

For the – due to the model – optimal point within this DoE, which had a $q_{s_{tot}}$ of 0.15, a specific titer of $0.16 \pm 0.05 \text{ g/g}$ is predicted. At a biomass concentration of 35 g/l this corresponds to a volumetric titer of $5.6 \pm 1.5 \text{ g/l}$. The low-low point (FB Nr. 9) had a specific titer of 0.18 g/g and a volumetric titer of 5.95 g/l , which is in-between of the predicted range.

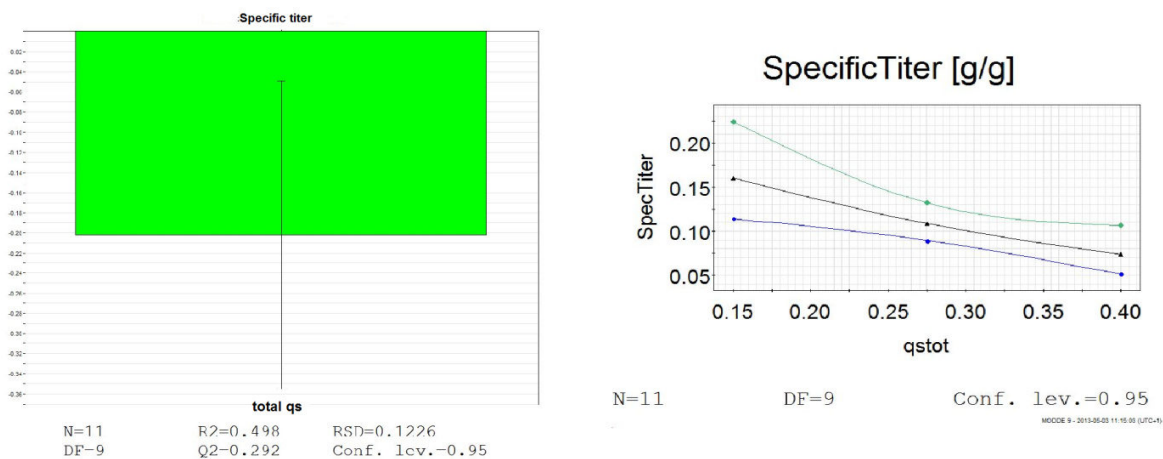


Figure 43: MLR model for the dependence of the specific titer on total specific substrate uptake rate

The MLR explains the variance in rhBMP-2 titer only by 50%. Hence there need to be other influences. Figure 44 shows the linear correlation between total q_s and specific titer. At low total q_s the data point spread more than at higher points. This goes in line with the observations in Figure 42 were at high total q_s the titer was lower, but for 0.25 and 0.15g/g/h no big differences occurred. On the right side in the before mentioned figure, the difference of specific titers dependent on $q_{s_{ara}}$ is showed. All three presented fermentations had a total q_s of 0.25, and only $q_{s_{ara}}$ varied. Due to the graphs in the right part of Figure 42 and Figure 44, $q_{s_{ara}}$ supposes to have a quadratic influence on the specific titer. This goes in line with investigations in this working group concerning the application of other products into the pBAD system. In the here conducted study this observation cannot be statistically confirmed, but it needs to be kept in mind for further investigations.

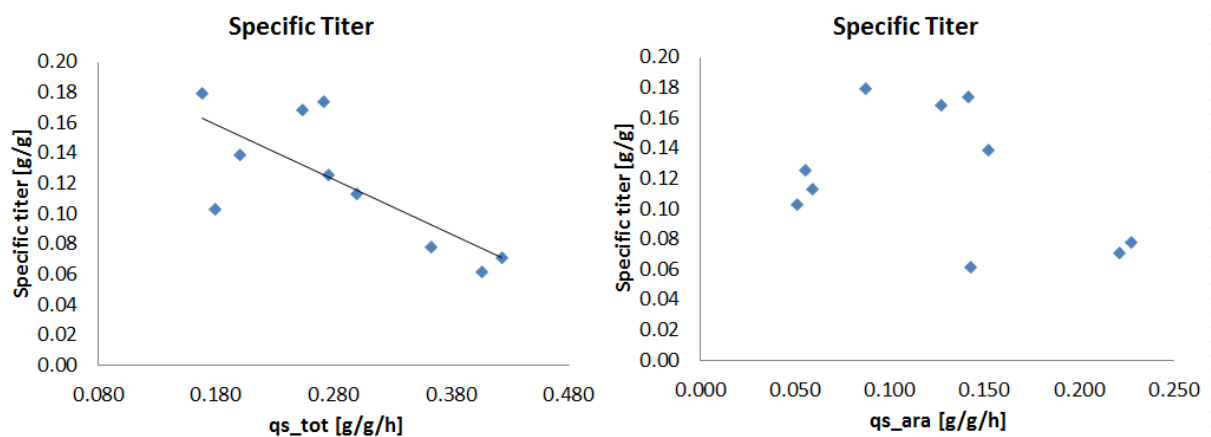


Figure 44: Connection between total specific substrate uptake rate and specific titer (left) and of the titer with specific L-arabinose uptake rate (right)

If fermentation 2 was excluded from calculations, the quadratic effect of $q_{s_{ara}}$ would become significant. In a model with $q_{s_{ara}}$ as only independent variable, this fermentation seems like an outlier. But it can be explained due to its high values of total specific substrate uptake rate.

The difference between specific titer and q_p is that the second mentioned is calculated out of two rates. Hence the error on this variable is higher than on the specific titer. This may be the cause of leaking significant correlation with other variables to explain its variability.

The findings for the process parameter influencing q_p (even if not significant) go in line with the findings of Sagmeister et al. [8]. The recombinant protein formation rate is solely dependent on the specific L-arabinose uptake rate, and not affected by the growth rate. An optimal point for

qp was not investigated in this study. The model is not good enough to make exact predictions. The quadratic effect of $q_{s_{ara}}$ can be explained by first the tunability of the pBAD system, and secondly by the metabolic load. At low $q_{s_{ara}}$ set-points, the level of qp increases until the optimal point is reached. This trend follows the triggering of recombinant protein formation rate by the amount of inducer (L-arabinose) fed [58]. From the optimal point qp decreases with further raising $q_{s_{ara}}$. This might be explained by the metabolic load of the bacterium [116]. Concerning the specific titer, only the total specific substrate uptake rate was found to have a significant impact. This might be linked to the longer induction periods at lower growth rates. Hence there is more time to build rhBMP-2. But the model only explains 50% of the variance in the specific rhBMP-2 titer. Hence there need to be a second variable having an influence. As shown above, $q_{s_{ara}}$ is proposed to have a quadratic impact, as it seems to have for the qp. This would make sense, as it can be expected, that an rhBMP-2 titer at optimal formation rates (triggered by $q_{s_{ara}}$) and low $q_{s_{tot}}$ (and hence long induction period), would reach its maximum.

3.3 Purity of rhBMP-2

The purity of rhBMP-2 was investigated *via* densitometric measurements on SDS gels. Beforehand bands representing the main impurities have been defined as reference bands, which were compared to the band at 13kDa, where the monomer of rhBMP-2 was expected. These reference bands were at 30, 37, 40, 49 and 60 kDa. Every sample was evaluated in triplicate. Figure 45 shows the three gels for fermentations 3 and 5. In the lanes loaded with homogenization supernatant and fermentation supernatant, no rhBMP-2 was found. The densitometric analysis was only conducted for samples of homogenization pellets (samples taken directly before lysis) and pellets from centrifugation after suspension in homogenization puffer, but without homogenization (samples taken after lysis).

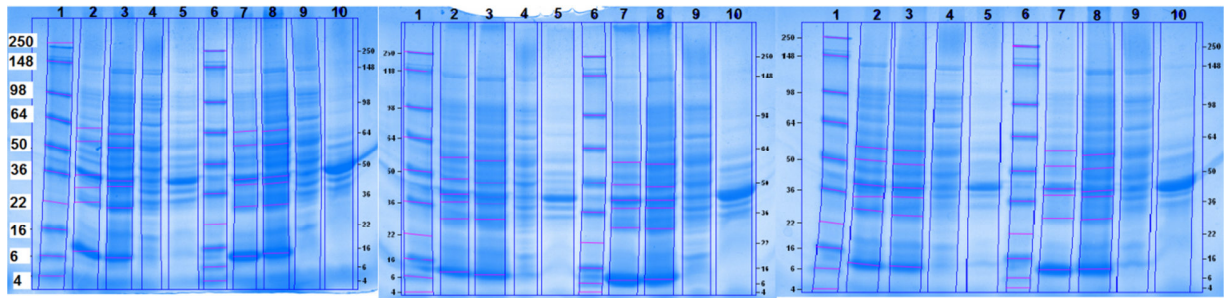


Figure 45: SDS gels pictures for fermentations 3 and 5. On all three gels the same samples were loaded in the same order. Lanes from left to right: 1&6: SeaBlue Plus 2 standard, 2: FB Nr. 3 pellet directly before lysis (homogenized), 3: FB Nr.3 pellet after lysis (not homogenized), 4: Homogenization supernatant FB Nr. 3 before lysis, 5: FB Nr.3 fermentation supernatant before lysis, 7: FB Nr.5 pellet directly before lysis (homogenized), 8: FB Nr.5 pellet after lysis (not homogenized), 9: Homogenization supernatant FB Nr. 5 before lysis, 10: FB Nr.5 fermentation supernatant

Despite the high measurement error in the densitometric evaluation of product purity (8%), a correlation between purity and total specific substrate uptake rate was found. The MLR model calculated for rhBMP-2 purity depending on total q_s had an R^2 of 0.521 and a Q^2 of 0.231 (see Figure 46). The regression was done excluding fermentation 7, as this run had been induced at too high biomass concentrations to be comparable. The prediction plot reflects the negative impact of total specific substrate uptake rate on purity. The highest purity of approximately 80% is predicted for total specific uptake rates smaller than 0.1g/g/h.

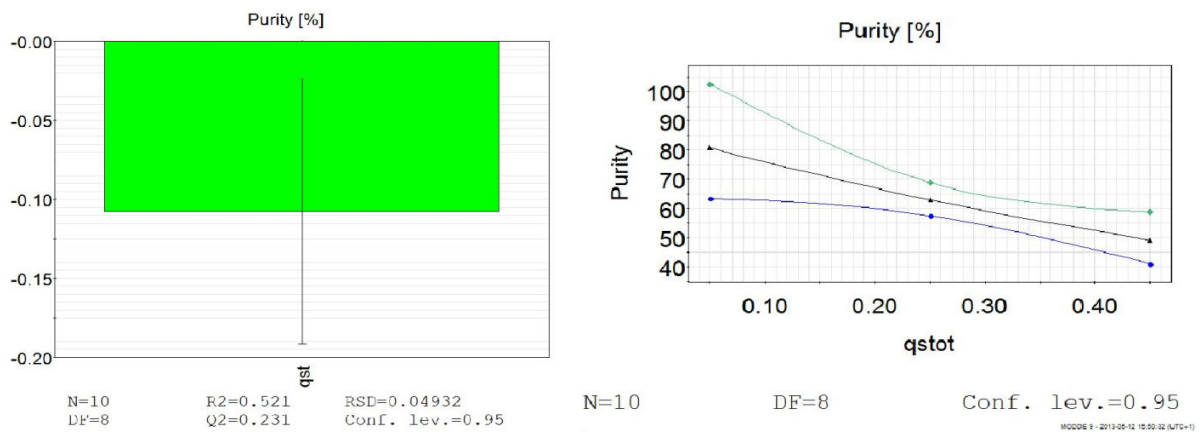


Figure 46: MLR for purity with total q_s as independent variable (left) and prediction of purity dependent on total q_s (right)

The densitometric method had a high experimental error, as the bands were marked by hand and also the width of each band was adjusted like this. Hence it is an operator-dependent method. Additionally the reference bands were defined at the beginning, and afterwards a band had to be marked at this defined protein size, even if no band was seen. Furthermore,

strong bands occurred at other protein sizes than the reference bands were defined, and thus not considered in the evaluation of purity.

On the other hand it was necessary to define bands beforehand to render comparability between the fermentations of DoE, but also with other samples including rhBMP-2 possible. Additionally it is an easy and cheap method to evaluate purity, and it was accurate enough to detect changes in purity. Due to previous investigations in this working group concerning temperature-dependency of purity, purity differences at 35°C induction temperature should be small compared to fermentations at 25°C. In the fermentations conducted here, purity varied between 49 and 79%, whereas the measurement error was 8%.

The model using total q_s as independent variable on purity, only explains 52% of variance. Furthermore, the Q^2 of 0.231 is too low to justify a good model, however the model is significant. Hence there need to be other parameters to have an impact on deviations in purity. Due to the PCA conducted in section 3.2.5 the specific glucose uptake rate could influence rhBMP-2 purity. But no effects were elicited in the here conducted study.

As already discussed in section 3.1.2, total specific substrate uptake rate, the growth rate and the induction period are co-linear variables. Hence these parameters ought to be exchangeable in a MLR model. Therefore the MLR was recalculated for induction period, including all eleven fermentations. Similar to the model presented for the total specific substrate uptake rate, the R^2 was 0.453 and the Q^2 0.228. It turned out that it was not possible to build a valid model when growth rate was chosen to be the independent variable.

The results of the negative impact of $q_{s_{tot}}$ on rhBMP-2 purity can be explained by the raising protein and ribosome level with increasing growth rates. At high growth rates, the transcription machinery for various proteins works faster than it does at lower growth rates [128]. Hence, the cytoplasmatic level of multiple proteins and ribosomes is lower at small total specific substrate uptake rates. Therefore the amount of impurities enclosed in the protein aggregates is rather low, when the growth rate is decreased.

3.4 Comparison of rhBMP-2 titer and purity in homogenized vs. lysed samples

In this chapter homogenized samples taken directly before lysis induction and samples with the highest lysis efficiency (not homogenized) are compared.

RhBMP-2 titer and time space yield

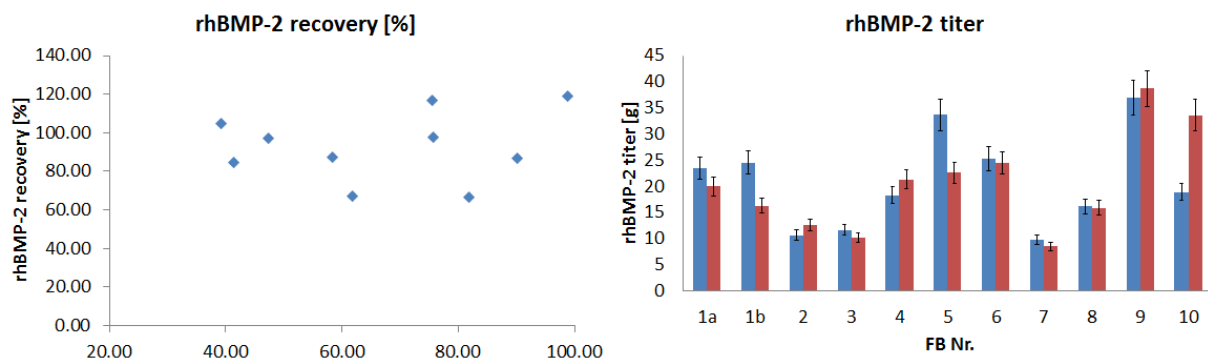


Figure 47: Correlation between rhBMP2 recovery and lysis efficiency (left) and rhBMP-2 titer before and after lysis

There was no connection between lysis efficiency and the amount of rhBMP-2 recovered after lysis (see Figure 47). The last sample before lysis induction had been homogenized and the amount of rhBMP-2 set as 100%. The sample with the highest lysis efficiency by way of comparison was not homogenized and its rhBMP-2 titer referred to the homogenized non-lysed sample. The amount of rhBMP-2 recovered after lysis in concern to the homogenized sample, was defined as the rhBMP-2 recovery.

When for all fermentations the homogenized sample taken before lysis was directly compared to the lysed sample, no significant differences in the rhBMP-2 titer were discovered (see Figure 47 of the right). This was approved by a paired t-test (20 ± 8.6 g before lysis, 20.4 ± 8.9 g after lysis, $p=0.4063$, $\alpha=0.05$). The H_0 "There is no difference in rhBMP-2 titer between homogenized, not lysed and lysed, not homogenized samples" was not rejected in favor of the alternative hypothesis "There is a significant difference in rhBMP-2 titer between homogenized, not lysed and lysed, not homogenized samples".

In contrast to titer, the time space yield differed significantly between not lysed homogenized and lysed not homogenized samples (see Figure 48 left). The paired t-test conducted in this concern showed values of 0.07 ± 0.02 g/l/h for homogenized and 0.06 ± 0.02 g/l/h for not

homogenized samples. The test was significant with a p of 0.0119 and an α -level of 0.05. Therefore, the H_0 “The time space yield for the homogenized not lysed samples is the same or lower than the time space yield for lysed not homogenized samples” was rejected. The alternative hypothesis was “The time space yield for the homogenized not lysed samples is the same or lower than the time space yield for lysed not homogenized samples” was rejected. The alternative hypothesis was “The time space yield for the homogenized not lysed samples is higher than the time space yield for lysed not homogenized samples”.

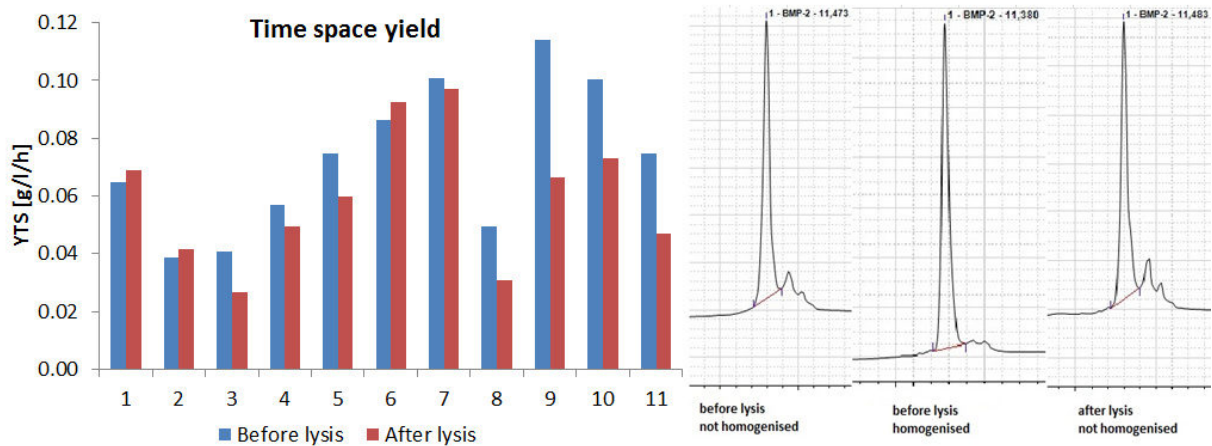


Figure 48: Comparison of times space yield (left) for not lysed, homogenized and lysed, not homogenized samples (left). Furthermore, HPLC peaks for lysed, not homogenized, for homogenized, not lysed and for lysed, not homogenized samples are shown on the right.

Initially it was hypothesized that the lysis efficiency should correlate with the rhBMP-2 recovery because the more cells E-lyse the more product should be able to leave the cell. This was not approved, as any connection between lysis efficiency and rhBMP-2 recovery was found.

Therefore, homogenization and non-homogenization was compared independently of lysis. Two not lysis samples were analyzed in parallel, wherefrom one had been homogenized and the other not. Figure 48 on the right shows, that there is no difference in rhBMP-2 titer between samples with and without applied cell rupture method. This can be explained by the high concentration (6M) of guanidine hydrochloride in the solubilization buffer. This reagent seems to be strong enough to open the cells and to therefore also solubilize inclusion bodies in former closed cells.

This means that neither of the two cell rupture methods – neither homogenization nor E-lysis – is distinct for the rhBMP-2 titer, when guanidine hydrochloride is used to solubilize the protein,

because this reagent is strong enough to recover the maximal amount of recombinant protein. As opposed to this, there was a difference in time space yield between lysed and homogenized samples. This finding could simply be explained by the fact that for E-lysis, the fermentation process took three hours longer than for a conventional fermentation. As there was no difference in rhBMP-2 titer, the three hours more of fermentation induced the variance in time space yield. With this parameter, time and space only in the upstream process are compared. Hence the downstream process is not considered. To better compare rhBMP-2 titer between homogenization and E-lysis in concern of time investigated and working volume, the downstream process needed to be considered.

Purity

In concern of purity differences between not lysed, homogenized and lysed, not homogenized samples were found. Homogenized samples are significantly purer ($60\pm 9\%$) than non-homogenized samples ($52\pm 7\%$) as investigated with a paired t-test ($p=0.0015$; $\alpha=0.05$). The H_0 "There is no difference in purity between lysed, not homogenized and homogenized, not lysed samples" was rejected in favor of the alternative hypothesis "There is a difference in purity between lysed, not homogenized and homogenized, not lysed samples".

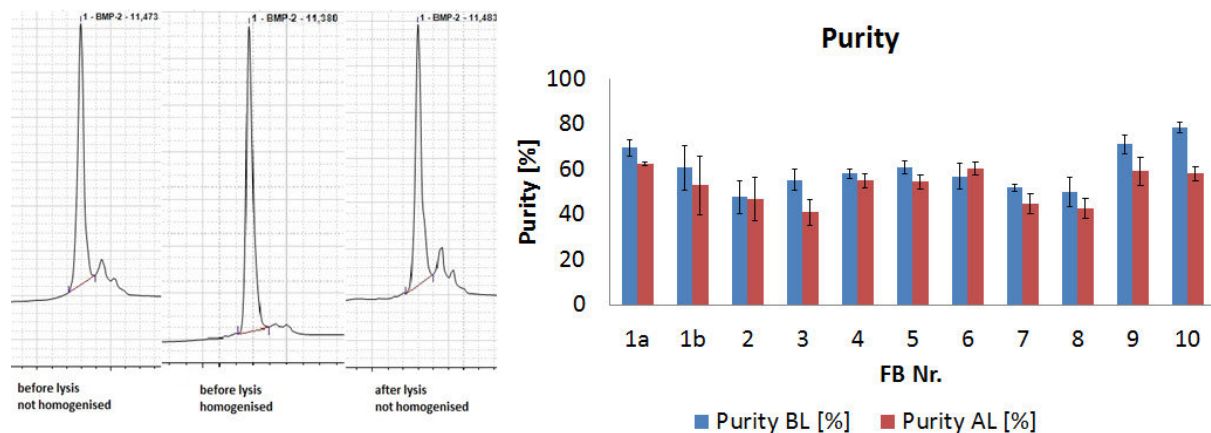


Figure 49: HPLC-peaks for various samples (left) and influence of homogenization and lysis on rhBMP-2 purity (right)

But homogenization still seems to be necessary when the product purity is regarded. As outlined above, the non-lysed homogenized samples showed to be purer than lysed non-homogenized samples. Additionally in the HPLC diagrams a second peak next to the one for

rhBMP-2 is found in non-homogenized samples (see Figure 49 on the left). There it was not distinguished whether a sample was lysed or not. The homogenization seemed to have the influence. Furthermore, the comparison of homogenization versus non-homogenization in not lysed cells *via* densitometric measurement on SDS gels showed that samples with homogenization are two times as pure as without (gel not shown here). This means, that with homogenization the purity can be doubled. But also after E-lysis samples are purer than without any cell rupture method, because in average the purity was only 10% smaller than after homogenization.

Homogenization may produce purer samples due to the fact that during homogenization the periplasmatic space is opened, whereas the soluble proteins of this compartment could be separated of inclusion bodies by centrifugation. Opposed to this non-homogenized non-lysed cells as well as BGs have the periplasmatic space closed. The compartment is opened during solubilization and its proteins found as impurities in the samples with the solubilized rhBMP-2. But when purities of homogenized and lysed samples are compared, only 10% of impurities could be explained by proteins from the periplasmatic space. There need to be other substances which by the opening of the cell in E-lysis or by the following centrifugation are separated from inclusion bodies.

Results of the chapter at a glance

The summarized findings of the comparisons between E-lysis and homogenization in bullet points:

- rhBMP-2 titer
 - No difference in rhBMP-2 titer between homogenized, not lysed and lysed, not homogenized samples
 - No difference in rhBMP-2 titer between homogenized, not lysed and not homogenized, not lysed samples
- Time space yield
 - Higher time space yield in homogenized, not lysed than lysed, not homogenized samples

- Purity
 - Higher purity in homogenized, not lysed than in lysed, homogenized samples
 - Higher purity in homogenized, not lysed and not homogenized, lysed than in not homogenized, not lysed samples

4. Conclusions

This study showed how the bacterial ghost technology was applicable on a fed-batch recombinant protein production process with pBAD as expression system. Thereby constraints for maximal rhBMP-2 titer and optimal rhBMP-2 purity were found, and separately the maximal lyses efficiency achieved. Furthermore, an intersection of pBAD and the BG technology was found, where through fusion of both systems (cell rupture replaced by BG technology) the highest rhBMP-2 titer was suspected.

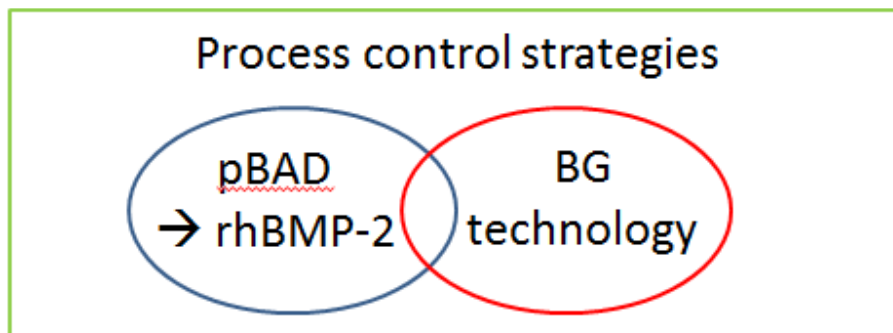


Figure 50: The fusion of the pBAD system to produce rhBMP-2 and of the BG technology. The study is framed by process control strategies in form of a first-principle soft-sensor

Additionally the first-principle soft-sensor was successfully applied on the control of CPPs in process development to investigate eleven points of a DoE.

4.1 Methodical conclusions

- **Applicability of first-principle soft-sensors for process development along QbD principles demonstrated**

Additionally to the findings of Sagmeister et al. (2013) concerning the dynamic process control *via* first-principle soft-sensors, first-principle soft sensors proved to be applicable for the control of specific rates as factors in DoEs. This tool guarantees a faster process development, because the yield coefficients need not to be investigated in advance. Additionally it is applicable even when yield coefficients vary over fermentation phases. Higher experimental precision and better models are provided as the set-points in a DoE are approached more accurately. Further the method is transferable to manufacturing processes, to ensure product quality.

- **Process control via first-principle soft-sensor may be distinct for process quality**

RhBMP-2 titer, product purity and lysis efficiency are all dependent on specific substrate uptake rate. Hence a control of these parameters by a first-principle soft-sensor can be important to achieve the goals of the process.

- **Multivariate data analysis**

Exploratory rate-based analysis using principal component analysis was used to examine interactions between process parameters and physiological variables. Two principal components were found to represent the influence of specific substrate uptake rate/specific growth rate (PC1) and the specific L-arabinose uptake rate (PC2) on the cell physiology.

The original CPPs specific D-glucose uptake rate and specific L-arabinose uptake rate were found to be transformable to total specific substrate uptake rate and percentage of L-arabinose in the feed. MVDA proofed to be an efficient tool for to describe an expression system and to investigate linkages between multiple variables in this system.

- **Lysis kinetics are describable by a negative exponential function**

Lysis kinetic was described *via* negative exponential functions with an average reaction constant of $-0.76 \pm 0.29\text{h}^{-1}$, basing on a differentiation of the culture into two sub-populations. When the parameters influencing lysis, namely specific substrate uptake rate and concentration of L-arabinose in the feed, are known, the examined reaction constant can be applied to calculate the period needed for E-lysis.

4.2 *Physiological conclusions*

- **Applicability of the bacterial ghost platform on recombinant protein production proved**

It was shown that E mediated lysis in *E. coli* high density fed-batch processes is possible up to a lysis efficiency higher than 95%. Furthermore, the influence of variances of parameters for the production of rhBMP-2 using the pBAD-system on the BG platform was investigated to predict lysis efficiency. To achieve lysis efficiency higher than 95% in a bioprocesses with a final biomass concentration of 35g/l, a minimal specific substrate uptake rate of 0.3g/g/h with a feed concentration of L-arabinose of 40% at the highest is necessary (see Figure 51)

- **Comparability of the Bacterial Ghost platform to conventional cell rupture methods**

Cell rupture via E-lysis was compared to homogenization. Concerning the product titer, no differences were found. In concerns of time space yield and product purity, homogenized samples showed significantly higher values. Purities after homogenization were in average $60 \pm 9\%$, after E-lysis $53 \pm 7\%$.

- **Highest product purity after homogenization, but lysed samples purer than samples without applied cell rupture method**

Through homogenization rhBMP-2 was purity doubled. E-lysed samples only had 10% more impurities than it was found in homogenized samples. Hence in concern of purity E-lysis is advantageous compared to waived cell rupture method, even when in this process the solubilization reagent guanidine hydrochloride was strong enough to disrupt the cells.

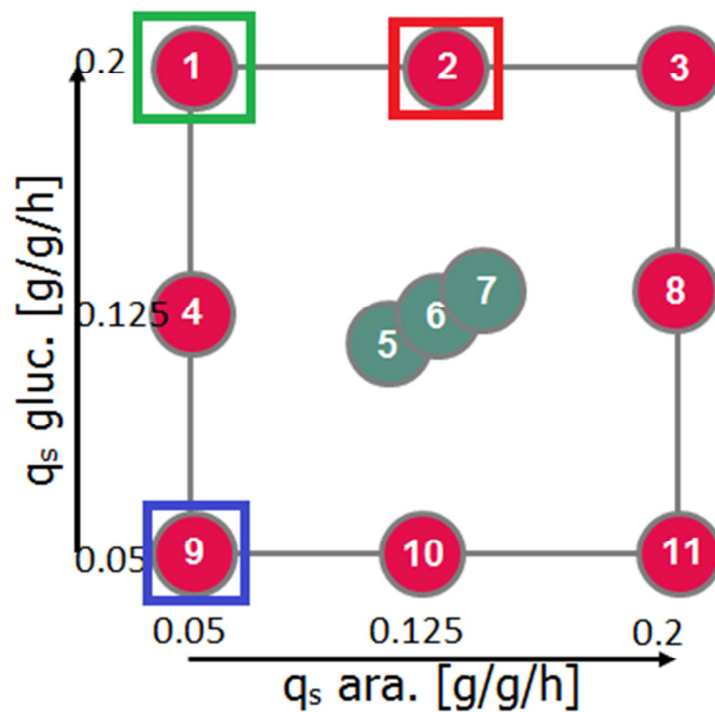


Figure 51: The optimal points inside of the DoE. The highest rhBMP-2 purity and the highest rhBMP-2 titer are achieved at the low-low point (marked in blue). Best lysis efficiencies are found at mean specific L-arabinose uptake rates and maximal total specific substrate uptake rates (marked in red). To achieve the highest rhBMP-2 titer when E-lysis is applied as only cell rupture method is found at 0.05g/g/h specific L-arabinose uptake rate and 0.2g/g/h specific D-glucose uptake rate.

- **Total substrate uptake rate as sole process parameters influencing specific rhBMP-2 titer: a negative influence**

Previous studies investigating the pBAD system applied on other products than rhBMP-2 have shown a negative influence of total specific substrate uptake rate as well as a

quadratic effect of specific L-arabinose uptake rate. In the here conducted study, only a negative influence of total specific substrate uptake rate appears with an optimum at a $q_{S_{total}}$ of 0.15g/g/h, whereas in addition a non-significant influence of specific L-arabinose is hypothesized. See Figure 51 for optimal point.

- **Product purity is negatively influenced by specific substrate uptake rates.**

The highest purity was achieved at a total specific uptake rate of 0.15g/g/h. Impurities can have an influence on further process steps like refolding of the protein. Hence higher impurities may bring along an additional purification step. Total specific substrate uptake rate of 0.15g/g/h is the optimal point in the here conducted DoE to achieve high rhBMP-2 titers and high product purity (see Figure 51).

- **The optimal point for the application of the Bacterial Ghost platform in fusion with the pBAD system to produce recombinant proteins is expected at 0.20g/g/h specific D-glucose uptake rate and 0.05g/g/h specific L-arabinose uptake rate. The expected final titer is 4 ± 0.84 g/l.**

The optimal point of rhBMP-2 titer was found to be at total specific substrate uptake rates of 0.15g/g/h. The expected volumetric titer for 35g/l biomass at this point is 5.6 ± 1.5 g/l. A lysis efficiency of more than 95% is reached at total specific substrate uptake rates of at least 0.3g/g/h and L-arabinose in the feed of 40%. Lysis efficiency has no impact on rhBMP-2 recovery. But if lysis efficiency correlated with the recovery of rhBMP-2, a compromise would have to be found were the combination of recombinant protein production, and its recovery out of the cells convey the highest amount of protein. High theoretical titers of protein are not gainful, when half of the protein rests in the cells. The specific titer at a total specific uptake rate of 0.3g/g/h (and 15% L-arabinose in the feed) is expected to be 3.85 ± 0.7 g/l, which at a lysis efficiency of 99% would give 3.81g/l proteins after lysis. For the optimal point concerning titer, a lysis efficiency of $53 \pm 13\%$ is expected (when the L-arabinose concentration is 33%). This would be a final rhBMP-2 of 3g/l (53% out of 5.6g/l). For a total specific uptake rate of

0.25g/g/h and 20% L-arabinose in the feed, a lysis efficiency of $92.5\pm 20\%$ is predicted. The volumetric titer is expected to be $4.55\pm 0.84\text{g/l}$, which at a lysis efficiency of 92.5% would promise a final titer of 4.2g/l. Hence, the point at a specific D-glucose uptake rate of 0.2g/g/h and a specific L-arabinose uptake rate of 0.05g/g/h ($0.25\text{g/g/h } q_{S_{\text{tot}}}$ and 20% L-arabinose in the feed, see Figure 51) is expected to be the optimal point for the fusion of the Bacterial Ghost technology and the pBAD system for the release of rhBMP-2. In the finding of the optimal point product purity was not considered, as before the impact of the here found impurities and the differences in purity needed to be further investigated, to interpret the importance of the outcome.

5. Outlook

- Extension of the investigated design space to receive higher product titers and higher purities

The highest titer as well as the highest purity of rhBMP-2 was received at the point with the minimal investigated total specific substrate uptake rate. The MLR model calculated out of process data suggests even higher titers and higher purities at lower specific substrate uptake rates. Hence the space of investigation needs to be extended to find the optimum of titer and purity.

- Investigation of the influence of impurities on diverse downstream process steps

As mentioned above, the influence of impurity is not known at this stage. In general impurities have a negative impact on protein titers after refolding [129]. Therefore it is necessary to investigate the impact of different impurities on the DSP as well as finding solutions to handle them.

- Investigation whether total specific substrate uptake rate/growth rate or the induction period are distinct for lysis efficiency

Due to the colinearity of total specific substrate uptake rate, growth rate and induction period, it needs to be investigated which of these parameters is distinct for lysis efficiency. This knowledge is necessary to predict lysis efficiency at constant total specific uptake rates, but different final biomass concentrations. If the only total specific substrate uptake rate was distinct, no difference in lysis efficiency should occur between final biomass concentrations of 35g/l or 45g/l. Also an interaction between total specific substrate uptake rate/growth rate and induction period is possible.

- Investigation of the inclusion body size and whether they leave the Bacterial Ghosts together with the cytoplasm

It is still not known whether the rhBMP-2 inclusion bodies are small enough to be released together with the cell content. If they could get out, the recombinant protein may be separated from the cell debris by for example gradient centrifugation. Further the solubilization of inclusion bodies would be separate from the cell debris, which would avoid opening the periplasm by the solubilizing reagent and further may avoid impurities coming from the periplasmic proteins. As alternative a reducing agent for inclusion bodies needs to be found, which does not disrupt cell walls of the bacterium. Then the inclusion bodies could be solubilized in a solution containing cell debris as well, without collecting impurities from the opening of periplasm. For the both described cases a correlation of lysis efficiency with recombinant protein recovery is expected, as the lysis tunnel would be distinct for the release for recombinant protein.

- Application of the Bacterial Ghost platform on other recombinant protein processes, especially soluble protein expression

This study has shown that the Bacterial Ghost technology can be fused with recombinant protein production processes. Unfortunately it was not possible to prove the affectivity of the

Bacterial Ghost platform for recombinant protein release out of the cytoplasm, because the solubilization reagent disrupted cells as well. Therefore the above described solutions could be applied. In the case of success, the application of the Bacterial Ghosts platform on other recombinant proteins as inclusion bodies could be investigated.

Still more promising seems the application of the Bacterial Ghost technology for recombinant proteins produced in a soluble form. After E-lysis, the protein could directly be separated from the cell debris by centrifugation.

- Fusion of Bacterial Ghost technology and recombinant protein excretion into the periplasm



After E-lysis the cytoplasmatic content is released into the medium, but the periplasmatic space is sealed. This could be used to separate the recombinant protein from the cytoplasmatic content including for example DNA and proteases, by centrifugation, where the medium and cytoplasm are separated from BGs containing the recombinant protein.

For the transport into the periplasm, the recombinant protein needs to be soluble, which does not allow high protein yields. Further the transport into the periplasm is expected to be the bottleneck in this process.

6. Appendix

6.1 Standard operation procedures

6.1.1 Protein precipitation

 TECHNISCHE UNIVERSITÄT WIEN VIENNA UNIVERSITY OF TECHNOLOGY	Standard Operating Procedure	
	Title	
<i>Research Division</i> Biochemical Engineering	SOP Number: SOP Status: Effective/Editable	Date:03.12.2012

Version	Number.0 for authorized versions; Number.1 for editions of a version, etc
Replaced version	
Author	Dominik Sauer
Date	03. 12. 2012
Signature	
Authorized by	
Date of authorization	
Signature	

Summary	TCA precipitation of proteins out of solubilisation buffer
----------------	--

Materials

- 10% w/v Trichloroacetic acid
- Aceton, cooled to -20°C
- 2 ml Eppendorf tubes
- 1 M Laemmli sample puffer – for SDS gel
- 0,1M/ 1% NaOH/SDS - for BCA measurement
-

Equipment

- Centrifuge

Procedure

- i) Prediction reagent preparation – 10% w/v Trichloroacetic acid : For example, 1g of TCA is dissolved in 10 ml deionized water and afterwards cooled.
 - ii) 1,5ml of the sample dissolved in the solubilisation buffer is centrifuged (10min, 16000g, room temperature). 1000 µl of the supernatant are transferred in a 2ml Eppendorf tube for further prediction. Your samples should show concentrations of 1-2mg/ml biomass dry weight otherwise dilute with deionised water.
 - iii) 1000µl of protein sample are mixed with 1000µl 10% w/v TCA solution and incubated 10min on 4°C for precipitation.
 - iv) After the samples are centrifuged (10min, 16000g, 4°C). The supernatant is discarded and the pellet of precipitate is washed with 1000µl -20°C acetone.
 - v) After a further centrifugation the acetone is discarded and the cap of the Eppendorf tubes is left open some minutes for evaporation of acetone.
 - vi) Sample preparation for SDS gel: the pellet of protein precipitate is dissolved in an adequate volume of 1M Laemmli sample buffer.
- Sample preparation for BCA measurement: the pellet of protein precipitate is dissolved in an adequate volume of 0,1M/ 1% NaOH/SDS buffer.
 -

If the precipitated protein is not completely dissolved in 0,1M/ 1% NaOH/SDS solution and the



solution is cloudy. Repeat the precipitation with a lower concentration of protein and wash twice with acetone.

Literature

-

End

6.1.2 Extracellular protein measurement via BCA kit

 TECHNISCHE UNIVERSITÄT WIEN VIENNA UNIVERSITY OF TECHNOLOGY	Standard Operating Procedure	
	Use of the Bicinchoninic Acid Kit for Protein Determination (Sigma product number BCA1-1KT)	
<i>Research Division</i> Biochemical Engineering		Date: 13.10.2010

Version	1.0
Replaced version	none
Author	Patrick Sagmeister
Date	13. 10. 2010

Summary	Use of the Bicinchoninic Acid Kit for Protein Determination (Sigma product number: BCA1-1KT)																		
Materials <ul style="list-style-type: none"> - BCA reagent A (Bichionic Acid Solution) - BCA reagent B (Copper II Solution) - Disposable plastic cuvettes 																			
Equipment <ul style="list-style-type: none"> - Spectrometer - Water bath 																			
Procedure <p>i) Working reagent preparation: 1ml of reagent will be used per sample and standards. Reagent is prepared by mixing 50 parts of <i>Reagent A</i> with one part of <i>Reagent B</i>. For example, for 51 tests 50ml of <i>Reagent A</i> and 1ml of <i>Reagent B</i> are mixed. The solution should appear light green in color.</p> <p>ii) Prepare a standard curve using 1ml BSA standard of 1mg BSA/ml. Use a 200μL piston pipette. As diluent, use the same buffer as used in your sample. Alternatively, BSA lyophilized standard can be used.</p> <table border="1" data-bbox="531 1261 1061 1534" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>conc [μg/ml]</th> <th>Standard [μl]</th> <th>Diluent [μL]</th> </tr> </thead> <tbody> <tr> <td>200</td> <td>40</td> <td>160</td> </tr> <tr> <td>400</td> <td>80</td> <td>120</td> </tr> <tr> <td>600</td> <td>120</td> <td>80</td> </tr> <tr> <td>800</td> <td>160</td> <td>40</td> </tr> <tr> <td>1000</td> <td>200</td> <td>0</td> </tr> </tbody> </table> <p>iii) Your samples should show concentrations in the calibration range. Otherwise dilute.</p> <p>For measurement of standards and samples, mix 50μl of protein sample with 1ml of the prepared BCA working reagent in an eppendorf tube and vortex. Choose one of the three possible incubation procedures:</p> <p>a) at 60°C using the water bath for 15 minutes.</p>		conc [μ g/ml]	Standard [μ l]	Diluent [μ L]	200	40	160	400	80	120	600	120	80	800	160	40	1000	200	0
conc [μ g/ml]	Standard [μ l]	Diluent [μ L]																	
200	40	160																	
400	80	120																	
600	120	80																	
800	160	40																	
1000	200	0																	

- b) at 37°C using the water bath for 30 minutes
- c) at room temperature for a minimum of 2 hours or overnight
- iv) After incubation, blank with working reagent and measure the absorbance of standards and samples at 562 nm.
- v) Use your standard curve to determine concentrations of the unknown samples.

Literature

- Sigma BCA Protein Assay Kit Technical Bulletin
- (Sigma product number: BCA1-1KT)

End

6.2 Abbreviations in formulas

6.2.1 Symbols

C_i	concentration of i	[g/l]
dM	mass change	[g]
$e_{H_2O, out}$	water fraction in off-gas	[mol/mol]
$F_{a, out}$	air outflow	[nl/min]
$F_{i, in}$	inflow of i	[g/h]
LE	lysis efficiency	[%]
\dot{M}_{IN}	mass input	[g/h]
\dot{M}_{OUT}	mass output	[g/h]
M_i	molar weight of i	[g/mol]
OD600	optical density at 600nm	[g/l]
p	probability value in statistical tests	[-]
q_{S_i}	specific conversion rate per gram of biomass	[Cmol/g/h]
r_i	conversion rate	[Cmol/l/h]
Q^2	goodness of prediction	[-]
R^2	goodness of prediction parameter	[-]
r_{inert}	inert gas ratio	[-]

RQ	respiratory quotient	[-]
SNR	signal to noise ration	[-]
t	expired time	[h]
V	volume	[l]
V_m	molar volume	[l/mol]
xAra	amount of L-arabinose in the feed	[%]
x_i	molar fraction	[-]
$Y_{m/i}$	yield C-mol m per C-mole of i	[-]
YTS	time space yield	[g/l/h]

6.2.2 Indices

aL	at the endpoint of lysis
ara	L-arabinose
b	base
bL	directly before lysis induction
i	component ID
in	into the fermenter
gluc	D-glucose
out	out of the fermenter
p	product (rhBMP-2)
s	substrate
tot	total substrate (L-arabinose + D-glucose)
x	biomass

6.2.3 Greek Symbols:

α	significance level	[-]
μ	specific growth rate	[1/h]
ρ	density	[g/l]

6.3 Data tables

Table 6: Comparison of offline measured and estimated biomass concentration at the end of induction phase

FB Nr. DA	Offline cx [g/l]	Estimated cx [g/l]
1a	29.5	35.2
1b	32.7	36.6
2	29.4	35.5
3	31.5	38.0
4	34.5	34.8
5	33.7	35.8
6	31.7	36.6
7	35.0	35.2
8	34.5	35.4
9	33.1	35.1
10	33.6	34.7

Table 7: Information matrix: Set-points and offline measured values for specific substrate uptake rates are shown.

Original Nr.	FB	DoE FB Nr.	Set-points				Measured values			
			qs_gluc [g/g/h]	qs_ara [g/g/h]	qs_tot [g/g/h]	xAra [%]	qs_gluc [g/g/h]	qs_ara [g/g/h]	qs_tot [g/g/h]	
31		1a	0.200	0.050	0.250	20.00	0.220	0.055	0.275	
38		1b	0.200	0.050	0.250	20.00	0.240	0.059	0.299	
27		2	0.200	0.125	0.325	38.46	0.263	0.143	0.406	
34		3	0.200	0.200	0.400	50.00	0.202	0.222	0.423	
26		4	0.125	0.050	0.175	28.57	0.119	0.051	0.179	
35		5	0.125	0.125	0.250	50.00	0.127	0.127	0.254	
36		6	0.125	0.125	0.250	50.00	0.131	0.141	0.272	
29		7	0.125	0.125	0.250	50.00	0.122	0.122	0.236	
30		8	0.125	0.200	0.325	61.54	0.135	0.227	0.362	
33		9	0.075	0.075	0.150	50.00	0.080	0.087	0.168	
32		10	0.050	0.125	0.175	71.43	0.049	0.152	0.200	

Table 8: Information matrix: Yield coefficients, rates, induction period and the C-balance

DoE FB Nr.	Induction period [h]	μ [1/h]	$Y_{x/s}$ [c-mol/c-mol]	RQ [c-mol/mol]	$Y_{o2/s}$ [c-mol/c-mol]	$Y_{co2/s}$ [c-mol/c-mol]	$Y_{co2/x}$ [c-mol/c-mol]	C-balance
1a	9.52	0.08	0.34	1.09	0.42	0.46	1.90	0.83
1b	9.42	0.10	0.35	1.07	0.40	0.43	1.23	0.96
2	7.98	0.12	0.32	1.03	0.41	0.42	1.35	0.78
3	5.13	0.16	0.40	1.01	0.42	0.43	1.11	0.89
4	15.03	0.05	0.33	1.05	0.49	0.52	1.62	0.86
5	9.77	0.09	0.40	1.13	0.44	0.50	1.27	0.94
6	9.51	0.10	0.40	1.04	0.42	0.51	1.32	0.92
7	4.91	0.10	0.48	1.08	0.47	0.51	1.12	1.06
8	7.42	0.12	0.37	1.07	0.40	0.43	1.17	0.86
9	18.98	0.05	0.31	1.04	0.48	0.50	1.70	0.84
10	14.03	0.07	0.36	1.12	0.57	0.64	1.82	1.01

Table 9: Information matrix: rhBMP-2 titers and time space yield (YTS) in homogenized not lysed (bL) and not homogenized lysed (aL) samples

DoE FB Nr.	Volumetric Titer [g/l]	Specific Titer [g/g]	BMP total [g]	qp [g/g/h]	Titer bL [g]	Titer aL [g]	YTS bL [g/l*h]	YTS aL [g/l*h]
1a	3.71	0.13	19.14	0.09	23.55	19.98	0.07	0.06
1b	3.70	0.11	19.01	0.10	24.54	16.28	0.07	0.05
2	1.80	0.06	9.12	0.06	10.67	12.70	0.04	0.04
3	2.23	0.07	11.15	0.08	11.68	10.12	0.05	0.03
4	3.57	0.10	18.39	0.04	18.33	21.36	0.06	0.07
5	5.68	0.17	29.12	0.12	33.70	22.60	0.11	0.07
6	4.96	0.17	25.30	0.11	25.30	24.51	0.10	0.07
7	1.82	0.05	9.46	0.06	9.77	8.53	0.04	0.03
8	2.69	0.08	13.91	0.10	16.19	15.87	0.06	0.05
9	5.95	0.18	28.39	0.07	36.94	38.69	0.10	0.10
10	4.67	0.14	23.01	0.06	18.94	33.63	0.09	0.09

Table 10: Information matrix: Lysis efficiency (LE) before lysis induction (bL) and at the end of lysis (aL), as well as rhBMP-2 recovery, extracellular protein (e.p.) titer as well as purities before and after lysis are shown.

DoE FB Nr.	LE bL [%]	Lysis kinetics k	LE aL [%]	rhBMP-2 recovery [%]	Titer [g/l]	e.p. Purity [%]	bL Purity aL [%]
1a	13.44	-0.97	41.45	84.84	1.77	69.99	62.05
1b	11.26	-0.54	81.93	66.35	1.83	60.68	52.94
2	8.00	-1.25	98.95	119.00	1.20	49.68	46.66
3	4.47	-0.52	90.24	86.65	1.39	55.37	41.01
4	9.90	-0.74	75.50	116.52	0.60	58.13	55.01
5	4.46	-0.47	61.68	67.08	1.70	60.96	54.40
6	8.10	-0.70	47.39	96.85	2.57	56.85	60.31
7	7.08	-0.65	58.39	87.28	1.65	52.51	46.38
8	5.69	-0.92	75.65	98.04	1.89	48.58	42.94
9	6.91	-0.97	39.26	104.75	1.71	71.07	59.11
10	5.89	-0.68	30.99	177.52	1.76	78.51	58.12

6.4 Acquired raw and processed data

In the following online and offline raw data, as well as processed data in form of rates, yield coefficients and specific rates are shown for all fermentations. Figure 52 shows the originally planned DoE and the DoE as it was finally conducted. The fermentations will be shown as numerated in the figure. Fermentation 4 has already been shown in section 3.2.

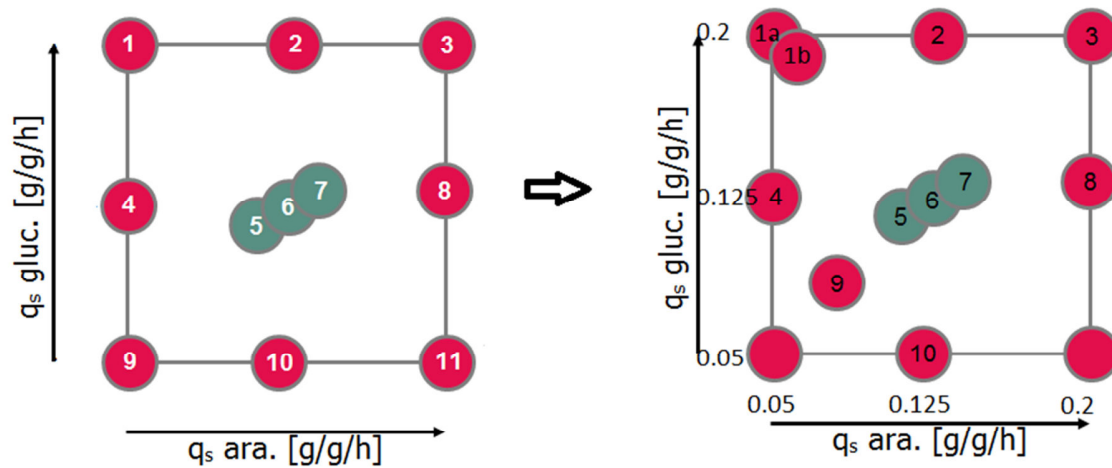


Figure 52: The planned set-points in the DoE (left) and the points finally conducted (right)

Online data of fermentation 1a:

Fermentation parameters: $q_{s_{gluc}}=0.2\text{g/g/h}$, $q_{s_{ara}}=0.05\text{g/g/h}$ (set-points)

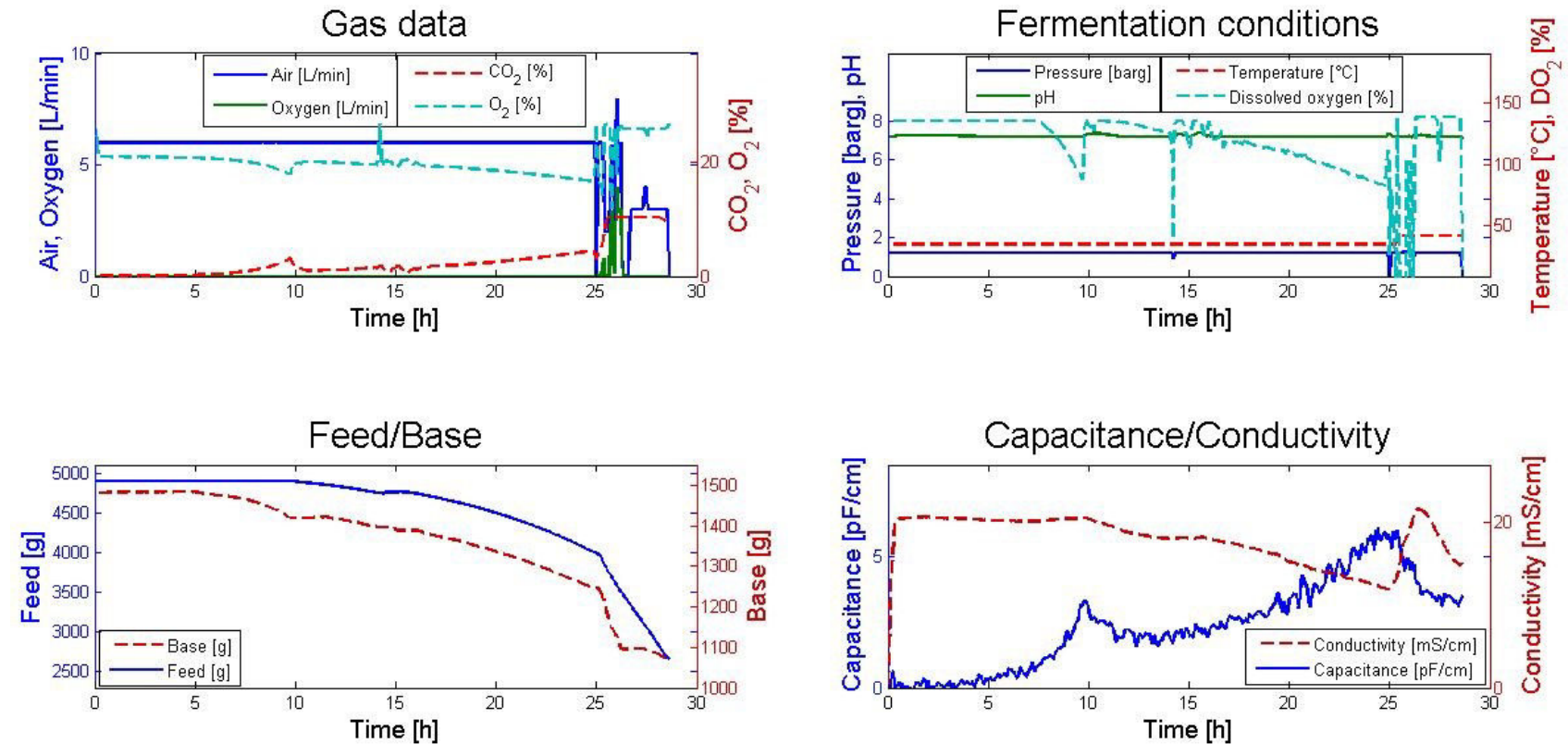


Figure 53: Online data of fermentation 1a. Gas data like air inflow (blue), oxygen inflow (green), CO_2 (red) and O_2 (cyan) in the off-gas are shown in the graph on the top left. On the top right fermenter pressure (blue), pH (green), fermentation temperature (red) and dissolved oxygen (cyan) are seen. In the bottom graphs feed (blue) and base (red) balance weights on the left and capacitance (blue) as well as conductivity (red) measurements on the right are presented.

Offline data of fermentation 1a:

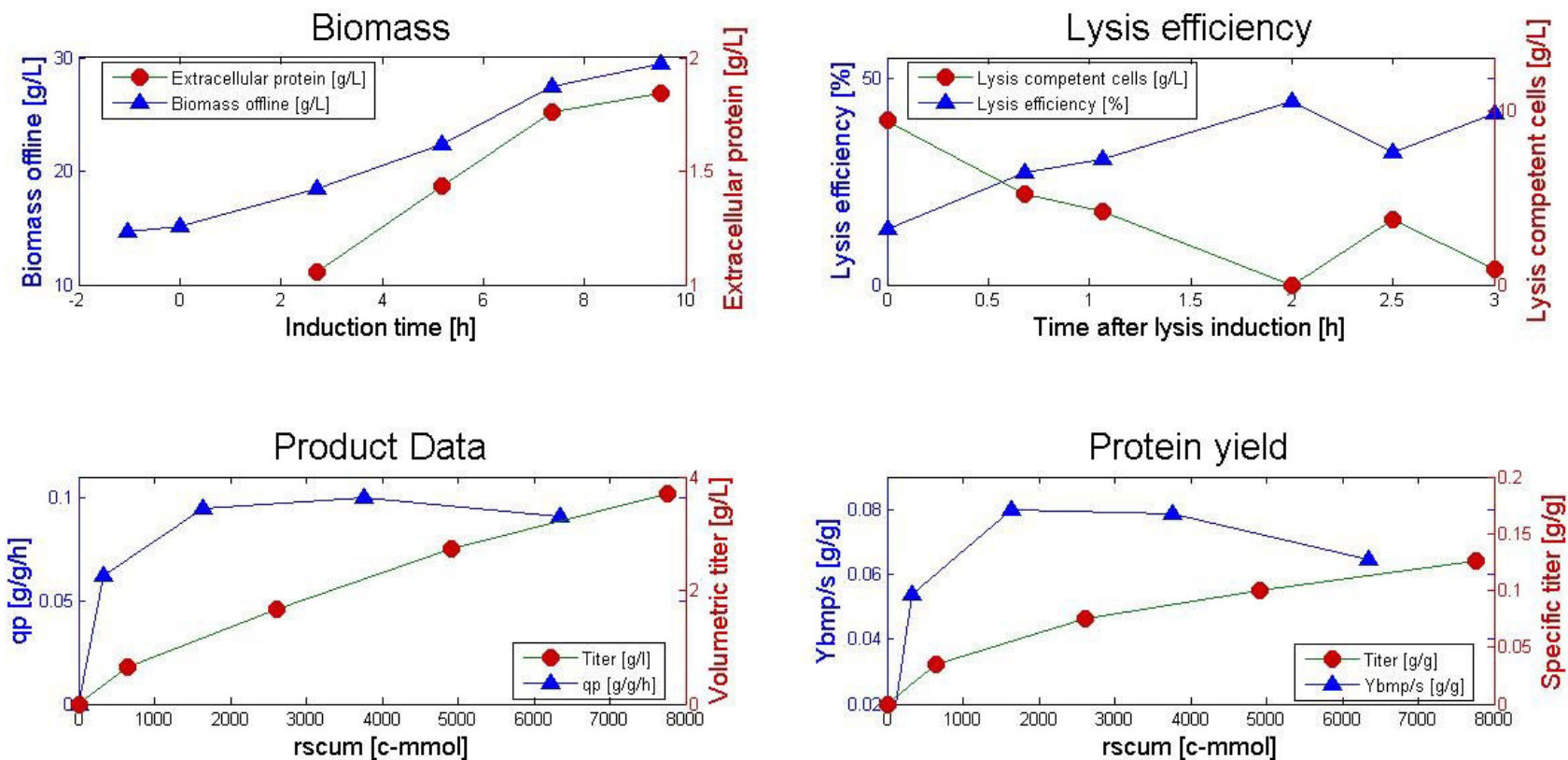


Figure 54: Offline data of fermentation 1a. On the top left biomass concentration (blue) and extracellular protein concentration (red) are shown over induction time. The beginning of mixed feed was defined as starting point. Lysis efficiency (blue) and lysis competent cells (red) are plotted in the graph on the top right. The graphs on the bottom present specific rhBMP-2 production rate (blue) and the volumetric rhBMP-2 titer (red) on the left as well as the rhBMP-2 yield (blue) and the specific rhBMP-2 titer (red) on the right. On the abscissa in the two last mentioned graphs, there is cumulated substrate. With one exception offline data is only shown for induction phase. Lysis efficiency and lysis competent cells are plotted for lysis phase.

Processed data of fermentation 1a:

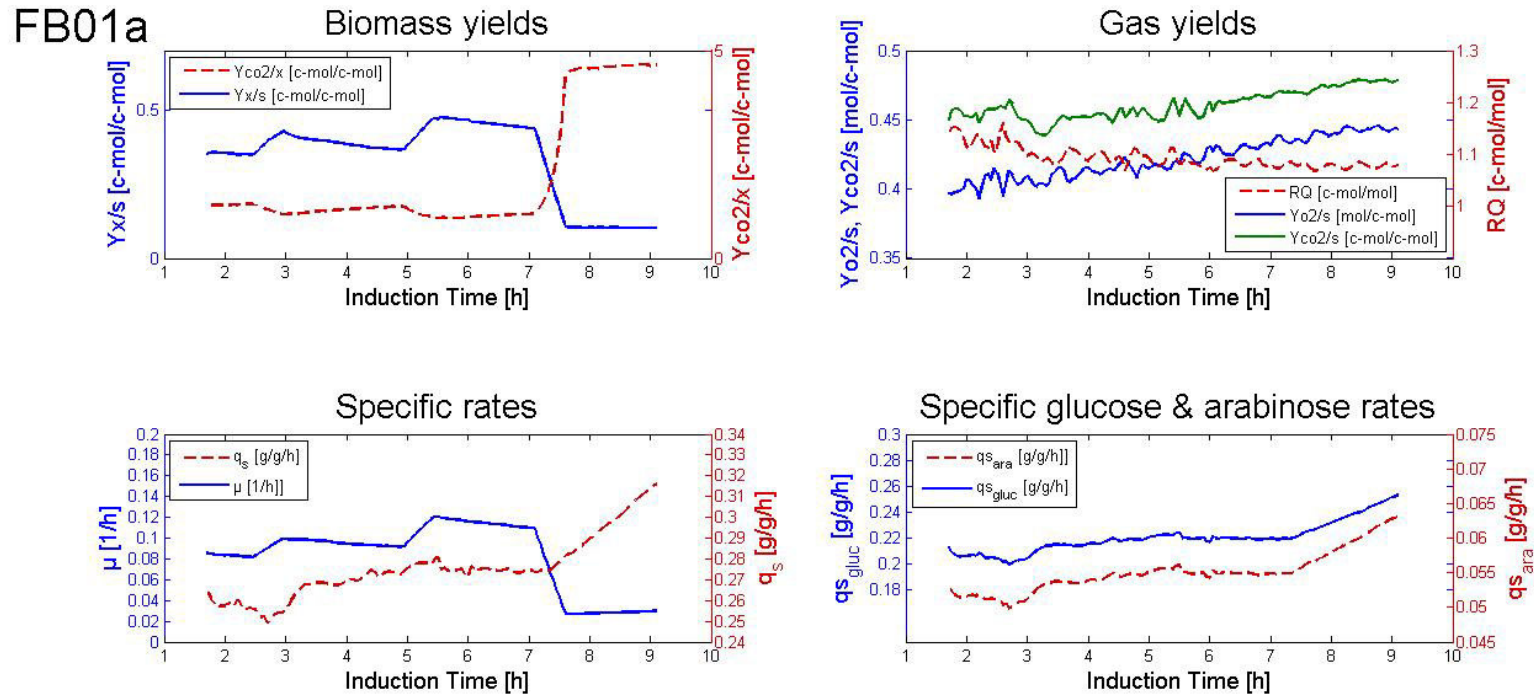


Figure 55: Processed data of fermentation 1a. Biomass yield (blue) and $Y_{CO_2/S}$ (red) are plotted in the graph on the top left. Next to it the gas yields are shown ($Y_{O_2/S}$ in blue, $Y_{CO_2/S}$ in green and the respiratory quotient in red). On the bottom the total specific substrate uptake rate (red) and the specific growth rate (blue) during recombinant protein production are found in the left, whereas specific D-glucose uptake rate (blue) and specific L-arabinose uptake rate (red) are on the bottom right. Processed data is only shown for the induction phase.

Online data of fermentation 1b:

Process parameters: $q_{s_{gluc}}=0.2\text{g/g/h}$, $q_{s_{ara}}=0.05\text{g/g/h}$ (set-points)

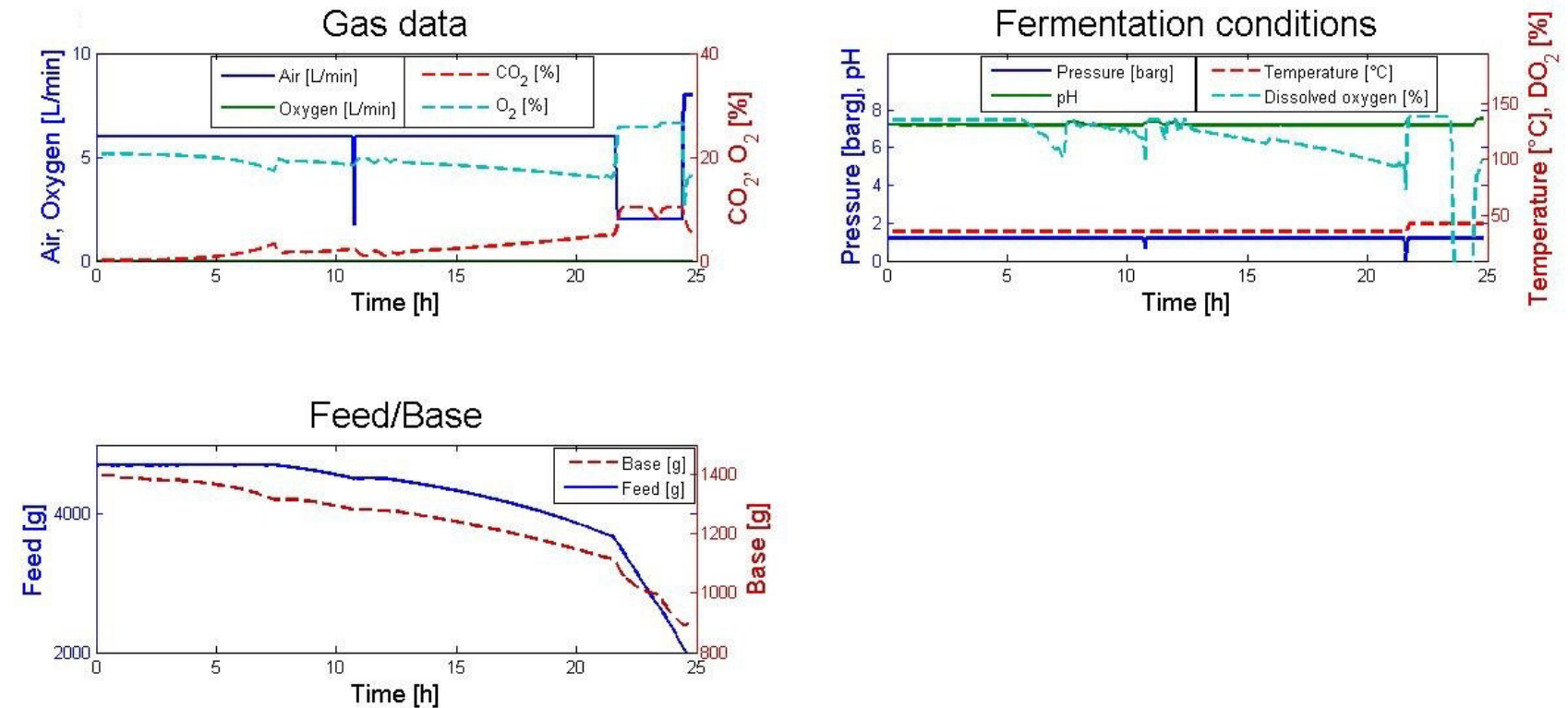


Figure 56: Online data of fermentation 1b. Gas data like air inflow (blue), oxygen inflow (green), CO₂ (red) and O₂ (cyan) in the off-gas are shown in the graph on the top left. On the top right fermenter pressure (blue), pH (green), fermentation temperature (red) and dissolved oxygen (cyan) are seen. In the bottom graphs feed (blue) and base (red) balance weights on the left are presented. Neither capacitance nor conductivity signals are shown.

In fermentation 1b the capacitance probe was not appropriately connected. Therefore neither capacitance nor conductivity measurements are available. The oxygen inflow signal in the graph is on zero for the whole fermentation. In lysis phase pure oxygen

was added, but the signals are not available, as the gas-flow was externally controlled. Only the decrease of air inflow signalizes the oxygen supplied.

Offline data of fermentation 1b:

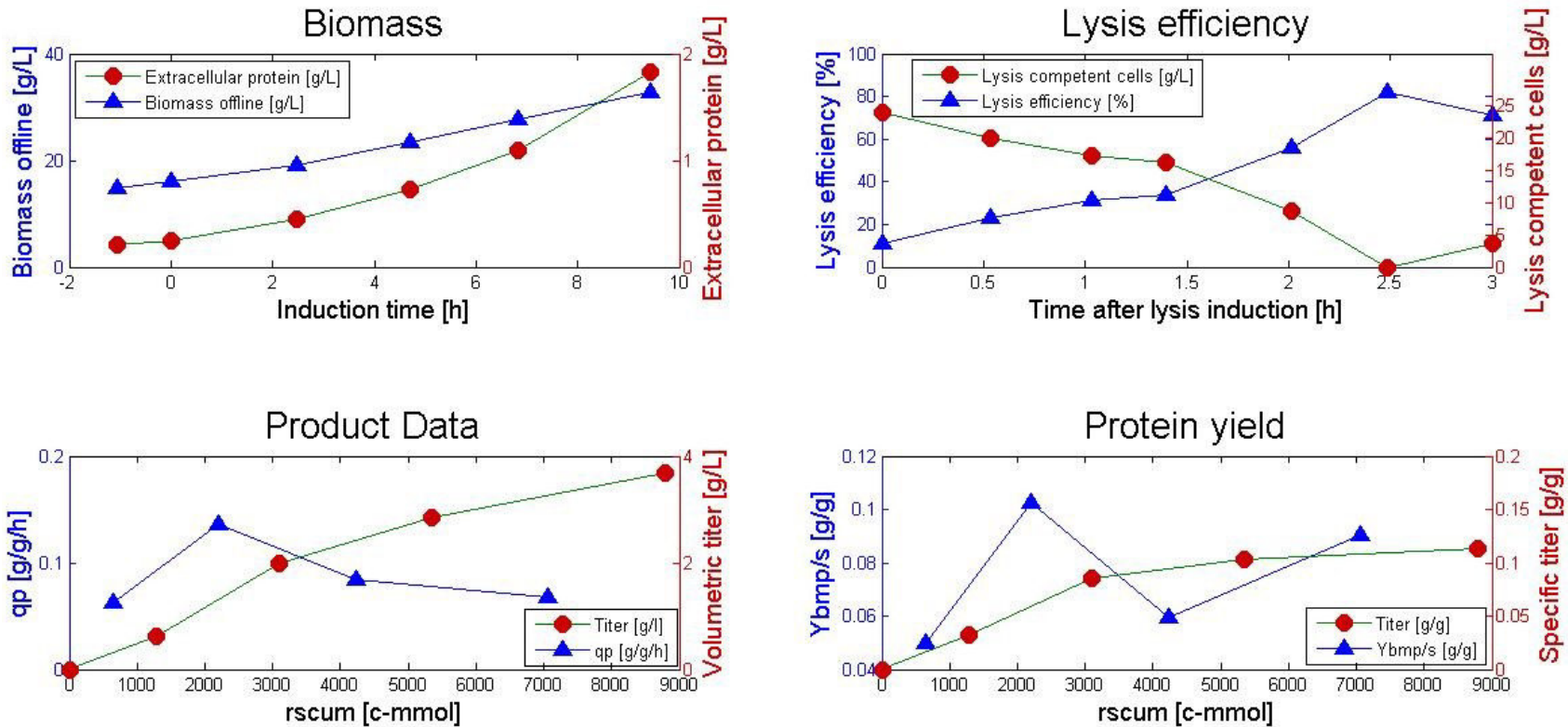


Figure 57: Offline data of fermentation 1b. On the top left biomass concentration (blue) and extracellular protein concentration (red) are shown over induction time. The beginning of mixed feed was defined as starting point. Lysis efficiency (blue) and lysis competent cells (red) are plotted in the graph on the top right. The graphs on the bottom present specific rhBMP-2 production rate (blue) and the volumetric rhBMP-2 titer (red) on the left as well as the rhBMP-2 yield (blue) and the specific rhBMP-2 titer (red) on the right. On the abscissa in the two last mentioned graphs, there is cumulated substrate. With one exception offline data is only shown for induction phase. Lysis efficiency and lysis competent cells are plotted for lysis phase.

Processed data of fermentation 1b:

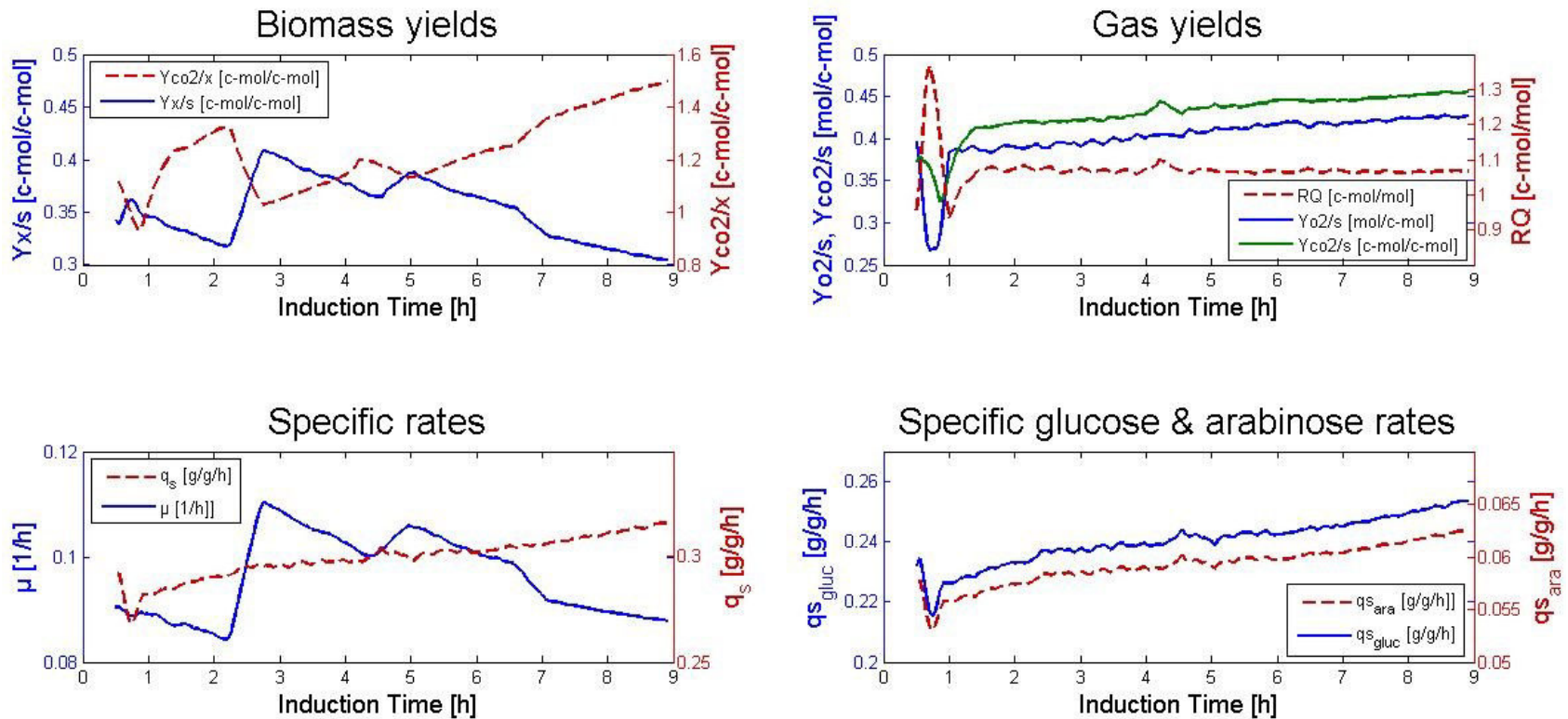


Figure 58: Processed data of fermentation 1b. Biomass yield (blue) and $Y_{co2/s}$ (red) are plotted in the graph on the top left. Next to it the gas yields are shown ($Y_{o2/s}$ in blue, $Y_{co2/s}$ in green and the respiratory quotient in red). On the bottom the total specific substrate uptake rate (red) and the specific growth rate (blue) during recombinant protein production are found in the left, whereas specific D-glucose uptake rate (blue) and specific L-arabinose uptake rate (red) are on the bottom right. Processed data is only shown for the induction phase.

Online data of fermentation 2

Process parameters: $q_{s_{gluc}}=0.2\text{g/g/h}$, $q_{s_{ara}}=0.125\text{g/g/h}$ (set-points)

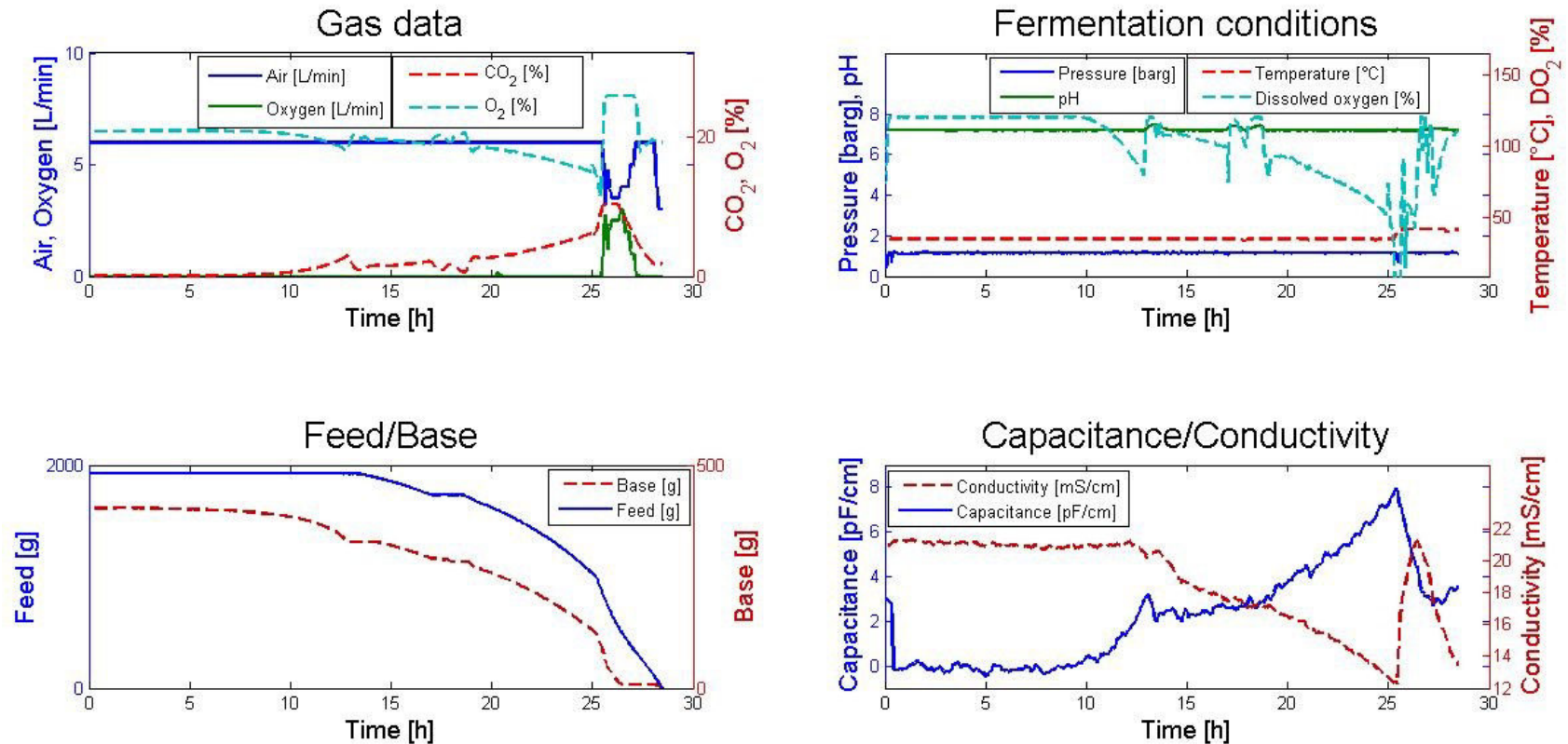


Figure 59: Online data of fermentation 2. Gas data like air inflow (blue), oxygen inflow (green), CO_2 (red) and O_2 (cyan) in the off-gas are shown in the graph on the top left. On the top right fermenter pressure (blue), pH (green), fermentation temperature (red) and dissolved oxygen (cyan) are seen. In the bottom graphs feed (blue) and base (red) balance weights on the left are presented as well as capacitance (blue) and conductivity (red) measurements on the right.

Offline data of fermentation 2:

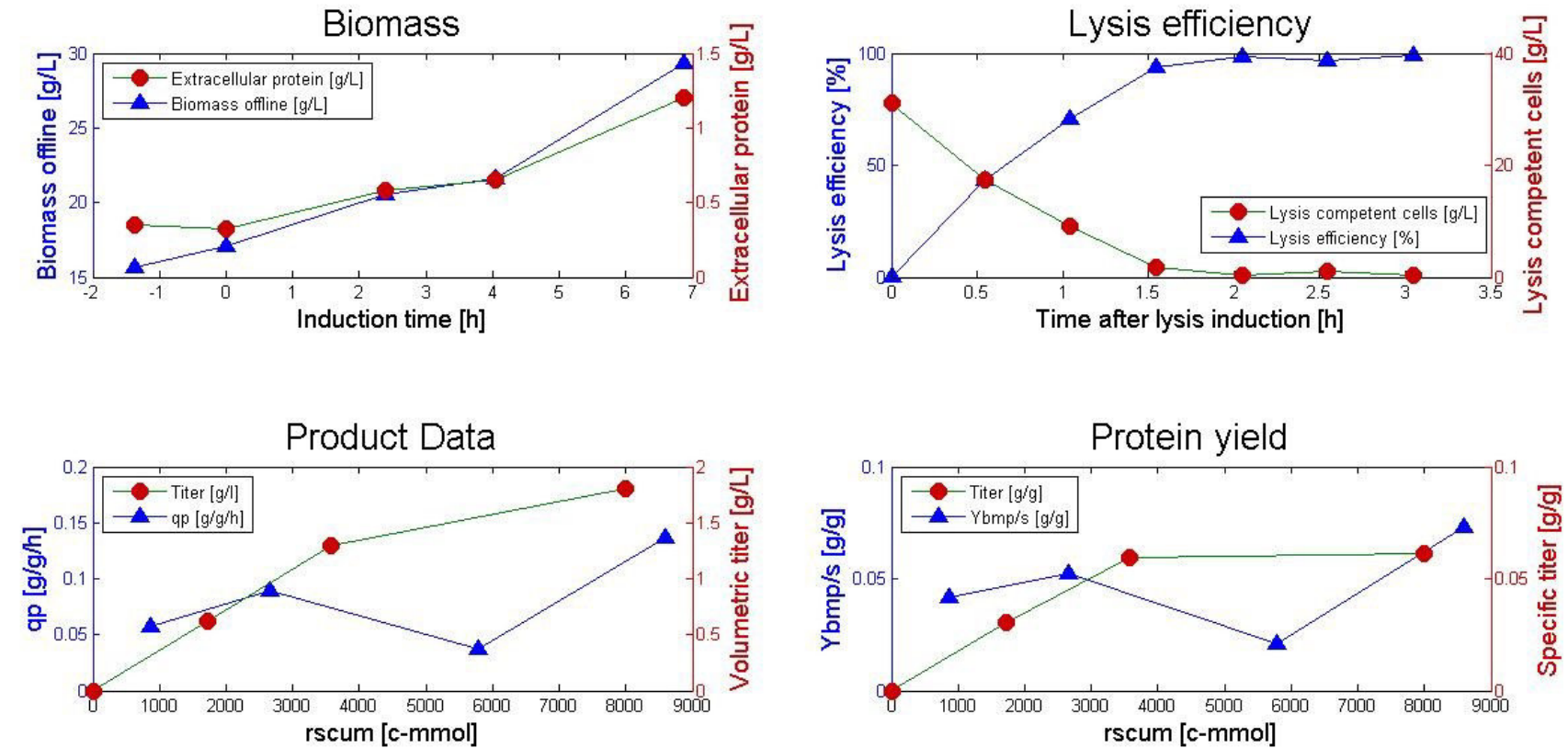


Figure 60: Offline data of fermentation 2. On the top left biomass concentration (blue) and extracellular protein concentration (red) are shown over induction time. The beginning of mixed feed was defined as starting point. Lysis efficiency (blue) and lysis competent cells (red) are plotted in the graph on the top right. The graphs on the bottom present specific rhBMP-2 production rate (blue) and the volumetric rhBMP-2 titer (red) on the left as well as the rhBMP-2 yield (blue) and the specific rhBMP-2 titer (red) on the right. On the abscissa in the two last mentioned graphs, there is cumulated substrate. With one exception offline data is only shown for induction phase. Lysis efficiency and lysis competent cells are plotted for lysis phase.

Processed data of fermentation 2:

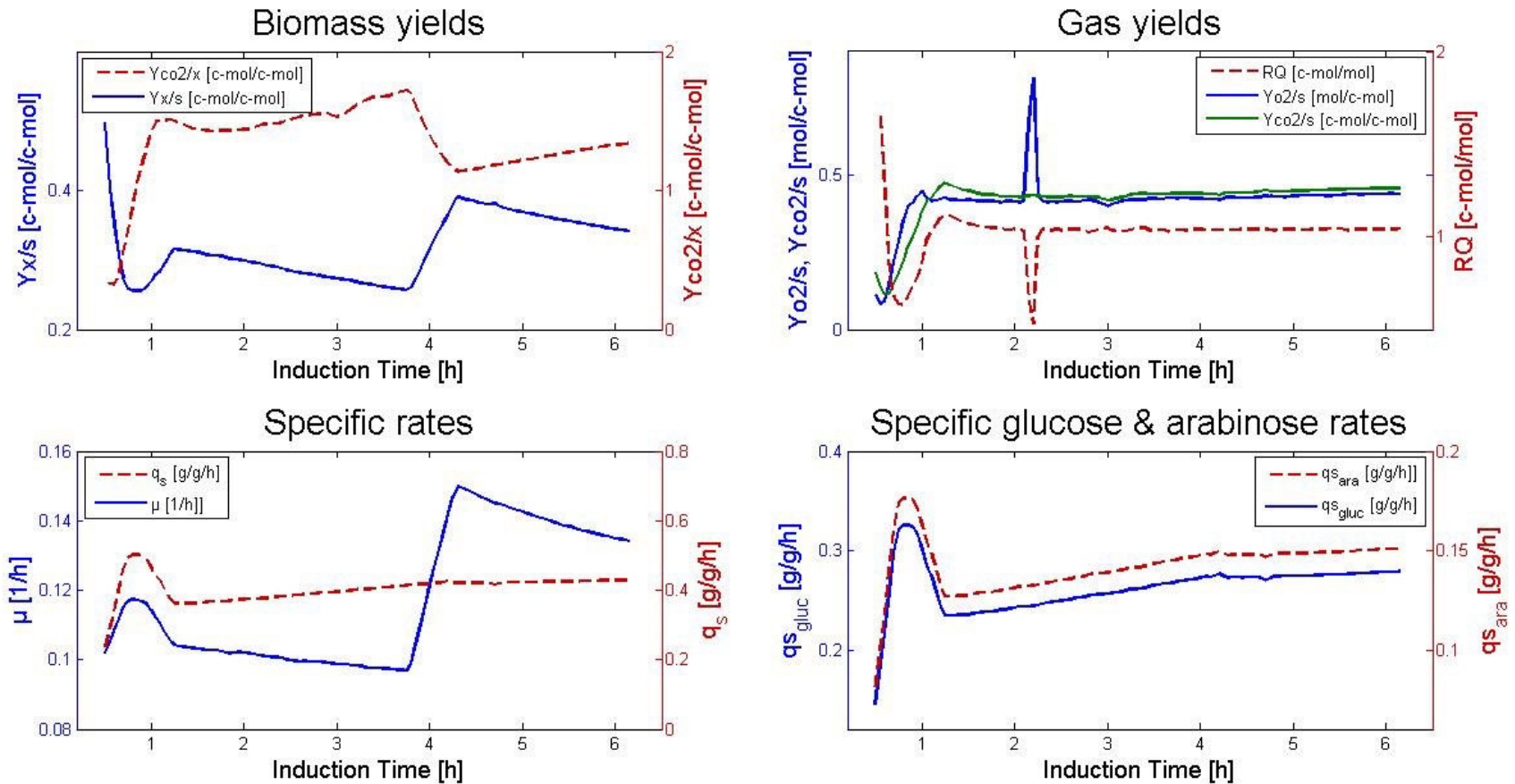


Figure 61: Processed data of fermentation 2. Biomass yield (blue) and $Y_{CO_2/s}$ (red) are plotted in the graph on the top left. Next to it the gas yields are shown ($Y_{O_2/s}$ in blue, $Y_{CO_2/s}$ in green and the respiratory quotient in red). On the bottom the total specific substrate uptake rate (red) and the specific growth rate (blue) during recombinant protein production are found in the left, whereas specific D-glucose uptake rate (blue) and specific L-arabinose uptake rate (red) are on the bottom right. Processed data is only shown for the induction phase.

Online data of fermentation 3:

Process parameters: $q_{s_{gluc}}=0.2\text{g/g/h}$, $q_{s_{ara}}=0.2\text{g/g/h}$ (set-points)

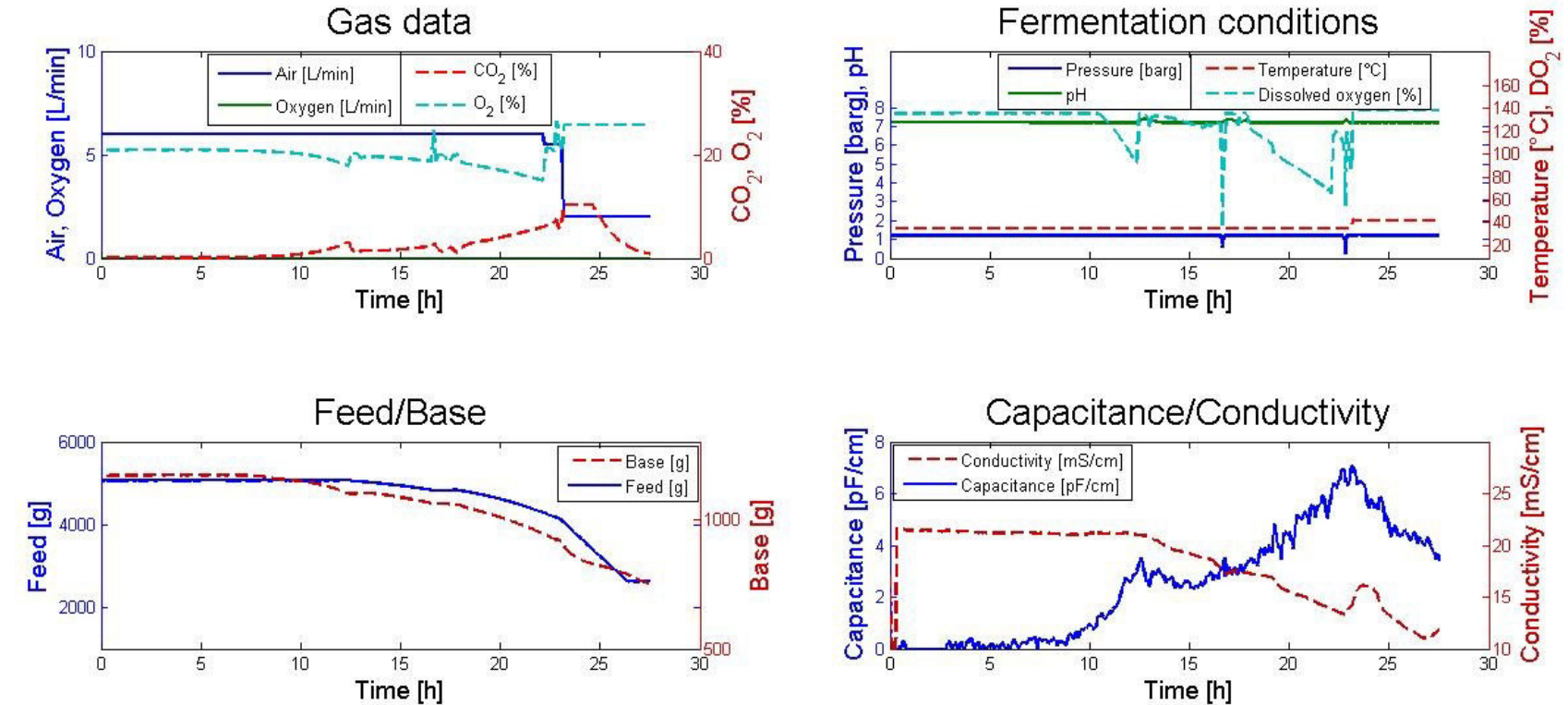


Figure 62: Online data of fermentation 3. Gas data like air inflow (blue), oxygen inflow (green), CO_2 (red) and O_2 (cyan) in the off-gas are shown in the graph on the top left. On the top right fermenter pressure (blue), pH (green), fermentation temperature (red) and dissolved oxygen (cyan) are seen. In the bottom graphs feed (blue) and base (red) balance weights on the left are presented as well as capacitance (blue) and conductivity (red) measurements on the right.

In fermentation 3 the pure oxygen supply was controlled with a mass flow controller of another fermenter. Therefore the real inflow is not shown in Figure 62.

Offline data of fermentation 3:

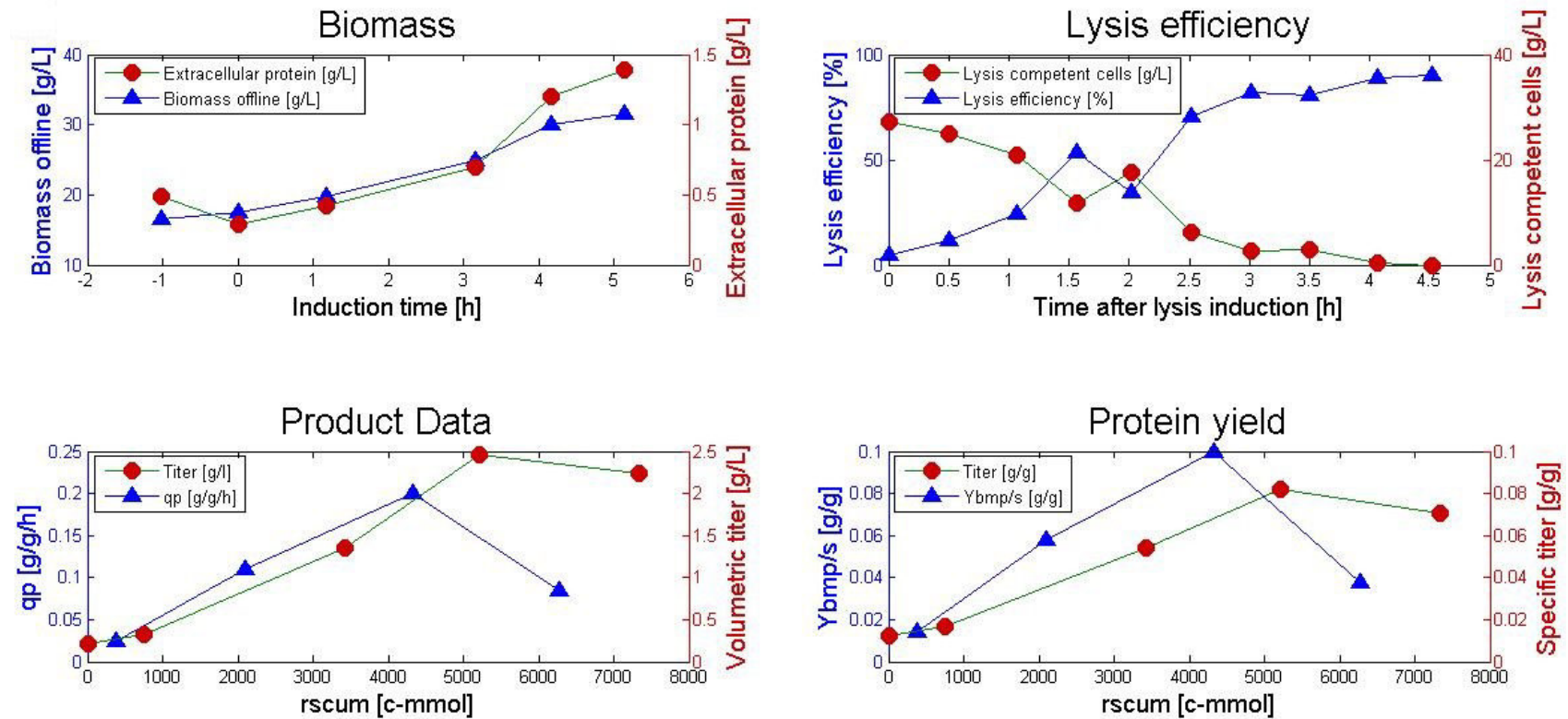


Figure 63: Offline data of fermentation 3. On the top left biomass concentration (blue) and extracellular protein concentration (red) are shown over induction time. The beginning of mixed feed was defined as starting point. Lysis efficiency (blue) and lysis competent cells (red) are plotted in the graph on the top right. The graphs on the bottom present specific rhBMP-2 production rate (blue) and the volumetric rhBMP-2 titer (red) on the left as well as the rhBMP-2 yield (blue) and the specific rhBMP-2 titer (red) on the right. On the abscissa in the two last mentioned graphs, there is cumulated substrate. With one exception offline data is only shown for induction phase. Lysis efficiency and lysis competent cells are plotted for lysis phase.

Processed data of fermentation 3:

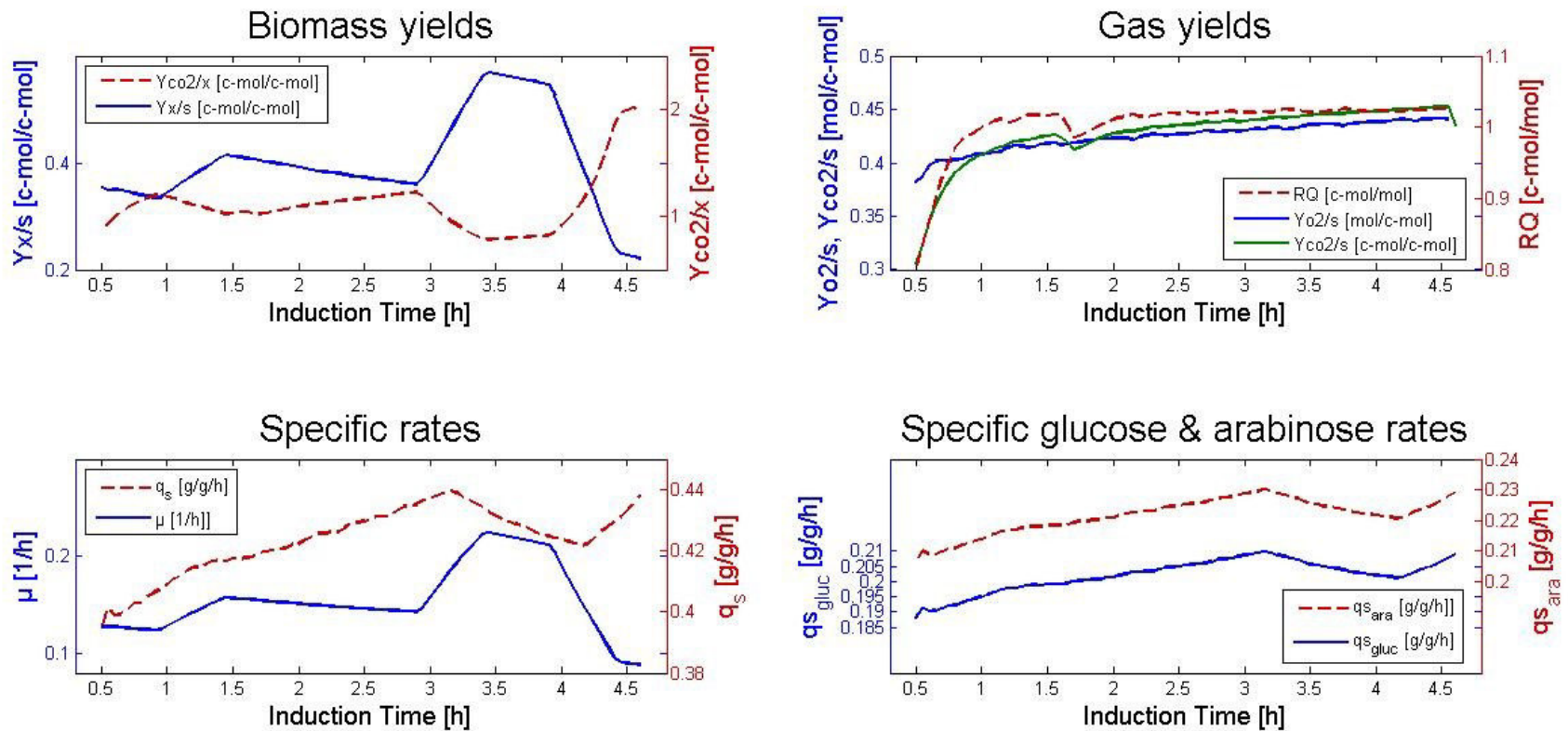


Figure 64: Processed data of fermentation 3. Biomass yield (blue) and $Y_{CO_2/s}$ (red) are plotted in the graph on the top left. Next to it the gas yields are shown ($Y_{O_2/s}$ in blue, $Y_{CO_2/s}$ in green and the respiratory quotient in red). On the bottom the total specific substrate uptake rate (red) and the specific growth rate (blue) during recombinant protein production are found in the left, whereas specific D-glucose uptake rate (blue) and specific L-arabinose uptake rate (red) are on the bottom right. Processed data is only shown for the induction phase.

Online data of fermentation 5:

Process parameters: $q_{s_{gluc}}=0.125\text{g/g/h}$, $q_{s_{ara}}=0.125\text{g/g/h}$ (set-points)

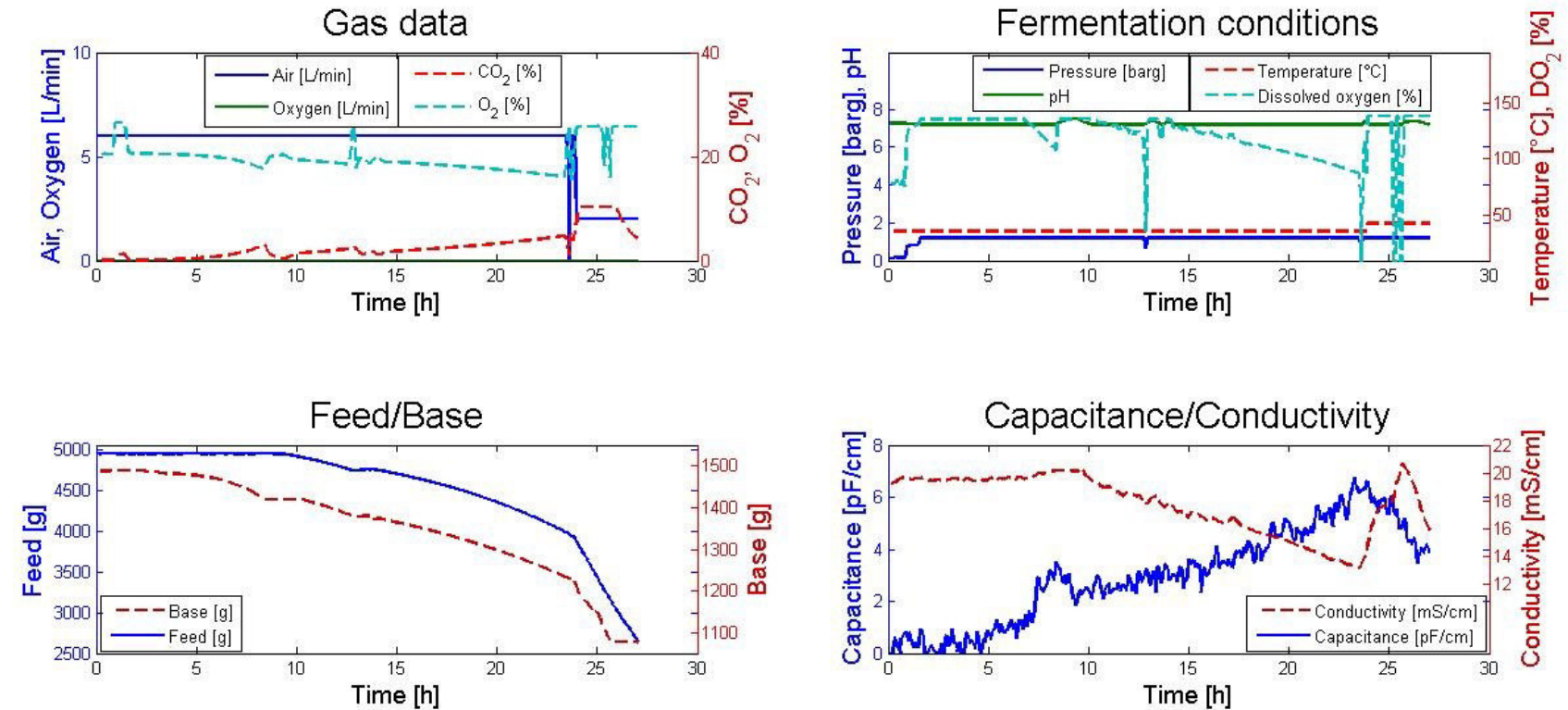


Figure 65: Online data of fermentation 5. Gas data like air inflow (blue), oxygen inflow (green), CO₂ (red) and O₂ (cyan) in the off-gas are shown in the graph on the top left. On the top right fermenter pressure (blue), pH (green), fermentation temperature (red) and dissolved oxygen (cyan) are seen. In the bottom graphs feed (blue) and base (red) balance weights on the left are presented as well as capacitance (blue) and conductivity (red) measurements on the right.

Again the oxygen inflow signal in Figure 65 is not representative for the real amount of oxygen added.

Offline data of fermentation 5:

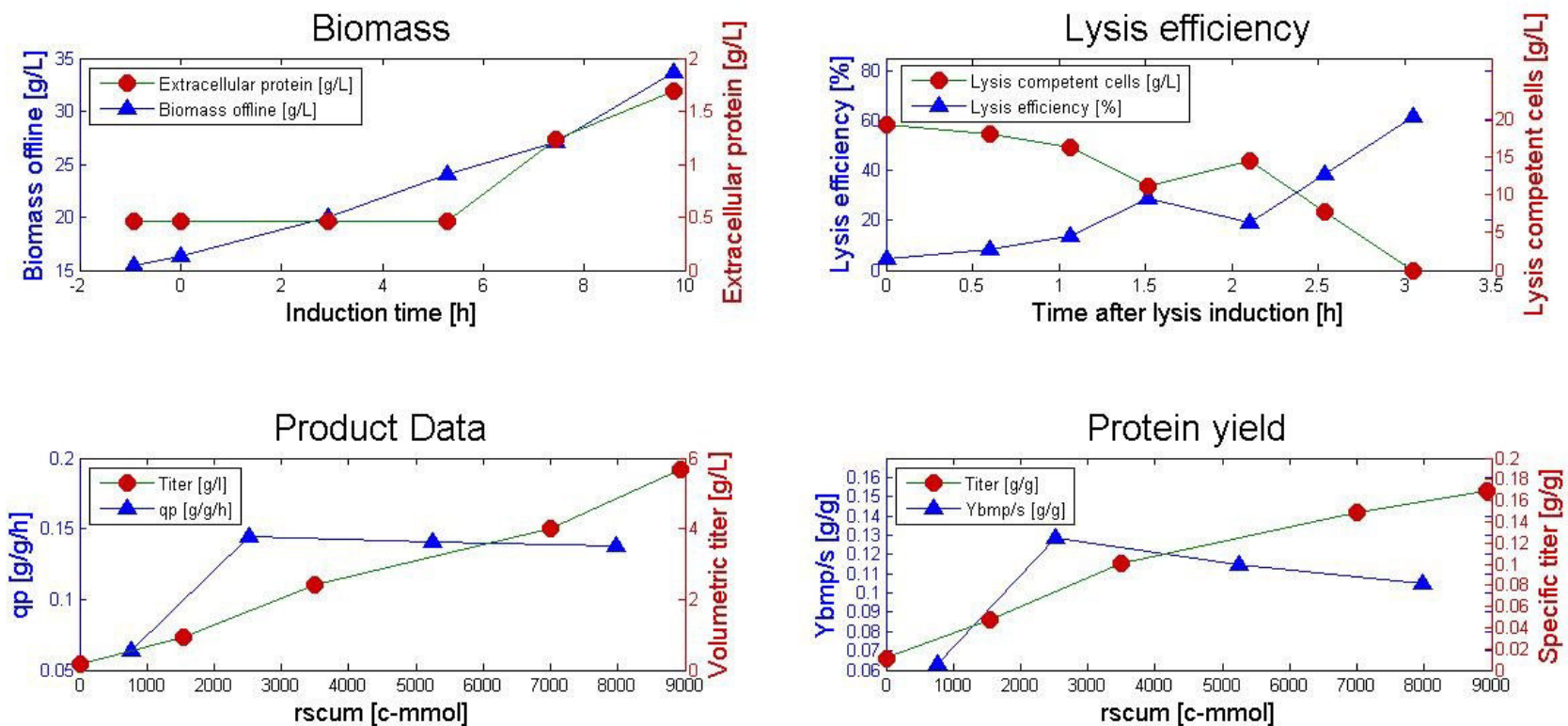


Figure 66: Offline data of fermentation 5. On the top left biomass concentration (blue) and extracellular protein concentration (red) are shown over induction time. The beginning of mixed feed was defined as starting point. Lysis efficiency (blue) and lysis competent cells (red) are plotted in the graph on the top right. The graphs on the bottom present specific rhBMP-2 production rate (blue) and the volumetric rhBMP-2 titer (red) on the left as well as the rhBMP-2 yield (blue) and the specific rhBMP-2 titer (red) on the right. On the abscissa in the two last mentioned graphs, there is cumulated substrate. With one exception offline data is only shown for induction phase. Lysis efficiency and lysis competent cells are plotted for lysis phase.

Processed data of fermentation 5:

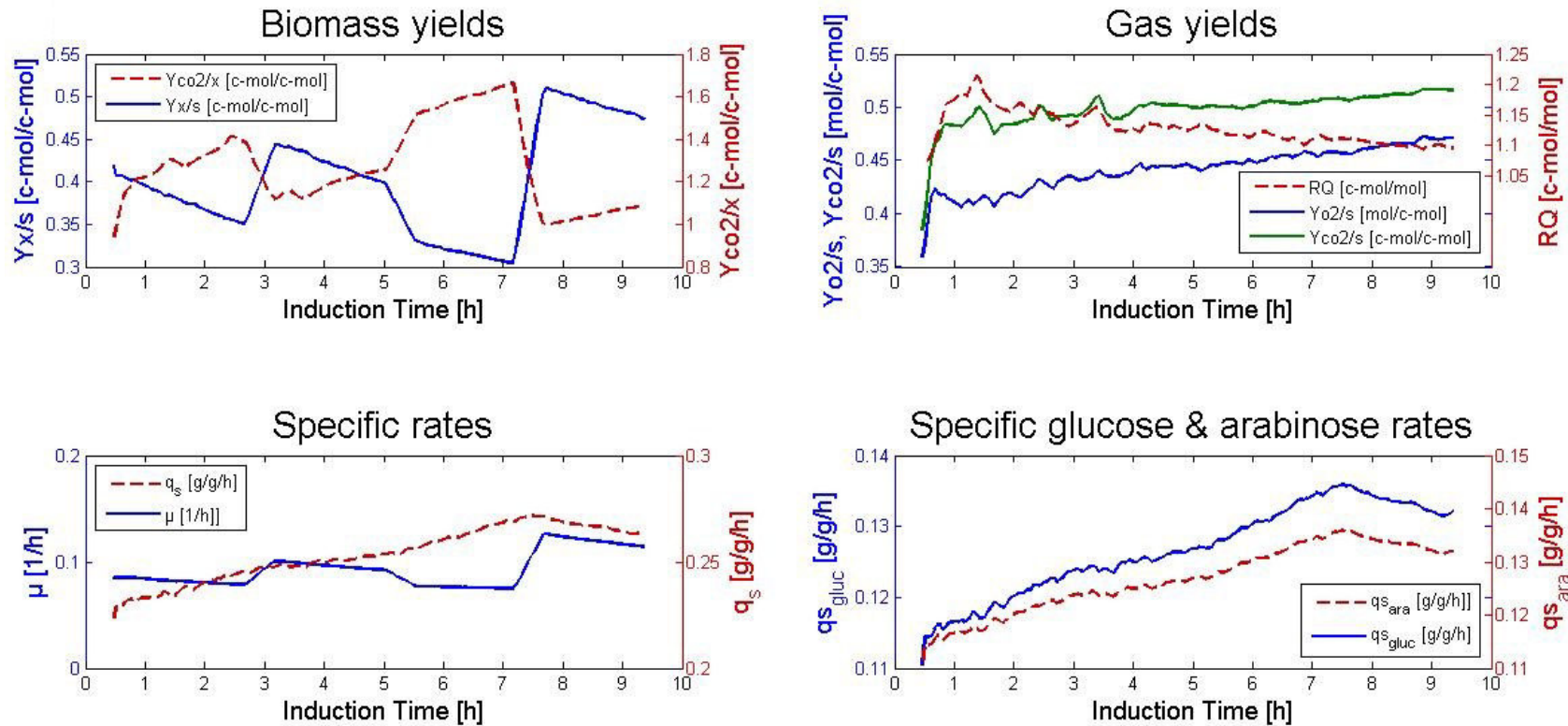


Figure 67: Processed data of fermentation 5. Biomass yield (blue) and $Y_{CO_2/s}$ (red) are plotted in the graph on the top left. Next to it the gas yields are shown ($Y_{O_2/s}$ in blue, $Y_{CO_2/s}$ in green and the respiratory quotient in red). On the bottom the total specific substrate uptake rate (red) and the specific growth rate (blue) during recombinant protein production are found in the left, whereas specific D-glucose uptake rate (blue) and specific L-arabinose uptake rate (red) are on the bottom right. Processed data is only shown for the induction phase.

Online data of fermentation 6:

Process parameters: $q_{s_{gluc}}=0.125\text{g/g/h}$, $q_{s_{ara}}=0.125\text{g/g/h}$ (set-points)

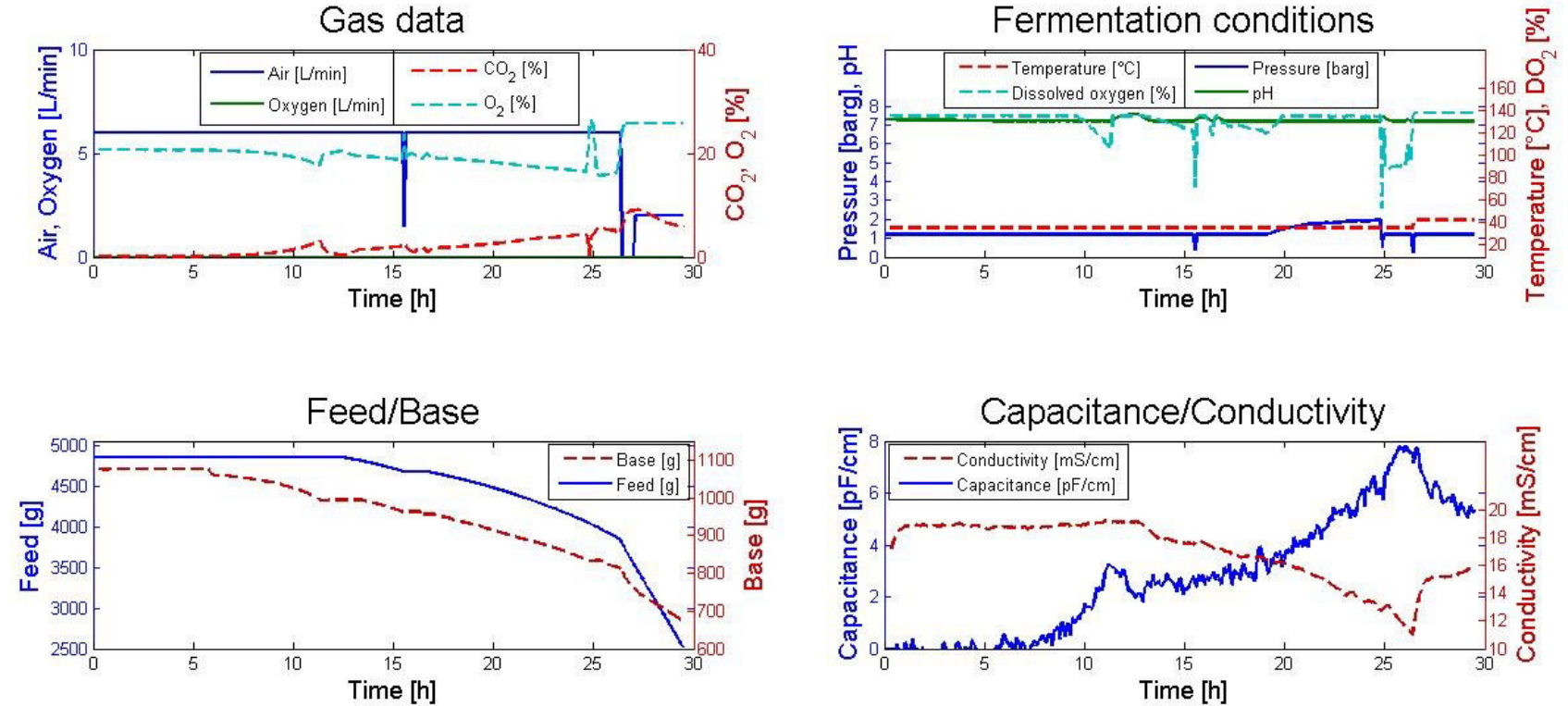


Figure 68: Online data of fermentation 6. Gas data like air inflow (blue), oxygen inflow (green), CO₂ (red) and O₂ (cyan) in the off-gas are shown in the graph on the top left. On the top right fermenter pressure (blue), pH (green), fermentation temperature (red) and dissolved oxygen (cyan) are seen. In the bottom graphs feed (blue) and base (red) balance weights on the left are presented as well as capacitance (blue) and conductivity (red) measurements on the right.

Not shown in Figure 68 is the oxygen supply during lysis phase, as it was controlled externally. The shifts in the air inflow signal represent the amount of oxygen added. The inflow of the gases summarized was always a total gas-inflow of 6l/min.

Offline data of fermentation 6:

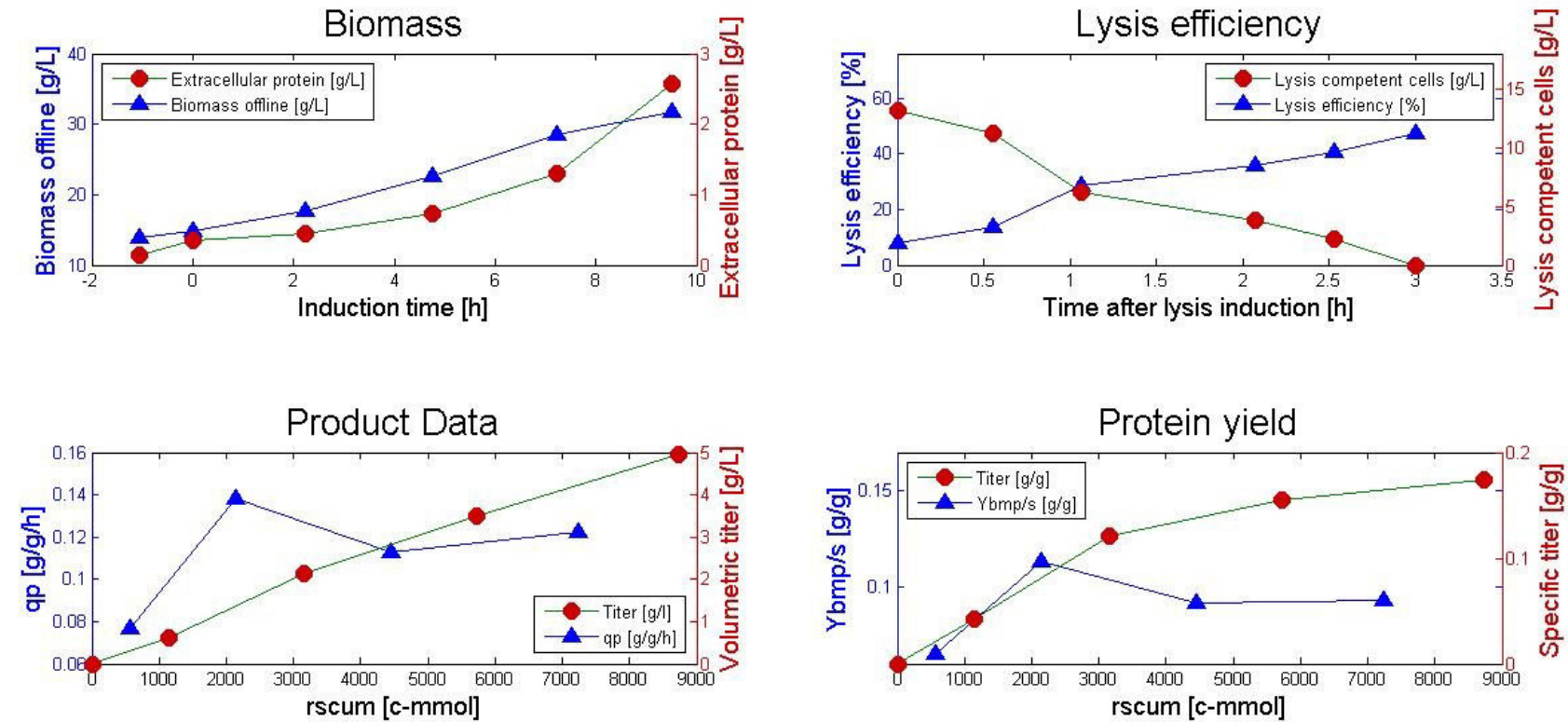


Figure 69: Offline data of fermentation 6. On the top left biomass concentration (blue) and extracellular protein concentration (red) are shown over induction time. The beginning of mixed feed was defined as starting point. Lysis efficiency (blue) and lysis competent cells (red) are plotted in the graph on the top right. The graphs on the bottom present specific rhBMP-2 production rate (blue) and the volumetric rhBMP-2 titer (red) on the left as well as the rhBMP-2 yield (blue) and the specific rhBMP-2 titer (red) on the right. On the abscissa in the two last mentioned graphs, there is cumulated substrate. With one exception offline data is only shown for induction phase. Lysis efficiency and lysis competent cells are plotted for lysis phase.

Processed data of fermentation 6:

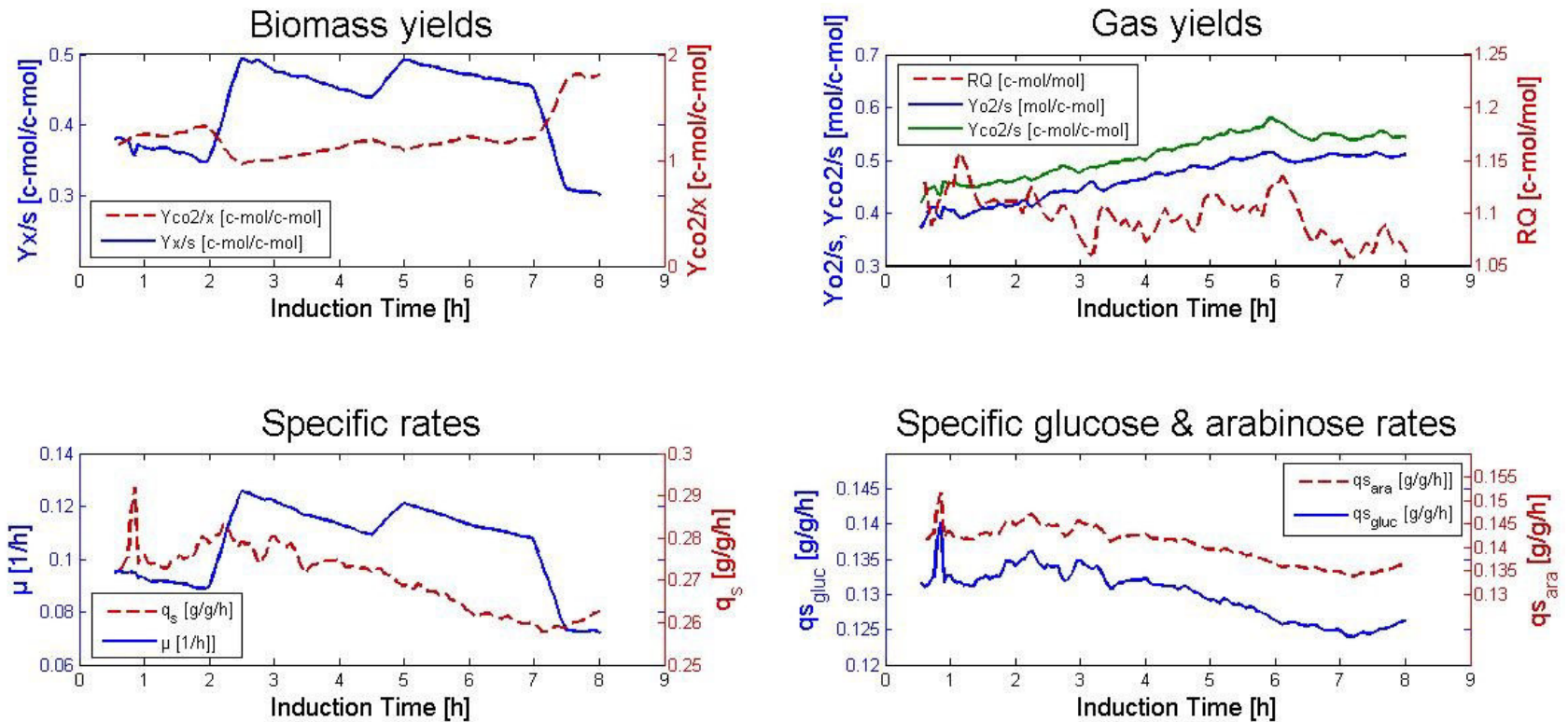


Figure 70: Processed data of fermentation 6. Biomass yield (blue) and $Y_{co2/s}$ (red) are plotted in the graph on the top left. Next to it the gas yields are shown ($Y_{o2/s}$ in blue, $Y_{co2/s}$ in green and the respiratory quotient in red). On the bottom the total specific substrate uptake rate (red) and the specific growth rate (blue) during recombinant protein production are found in the left, whereas specific D-glucose uptake rate (blue) and specific L-arabinose uptake rate (red) are on the bottom right. Processed data is only shown for the induction phase.

Online data of fermentation 7:

Process parameters: $q_{s_{gluc}}=0.125\text{g/g/h}$, $q_{s_{ara}}=0.125\text{g/g/h}$ (set-points)

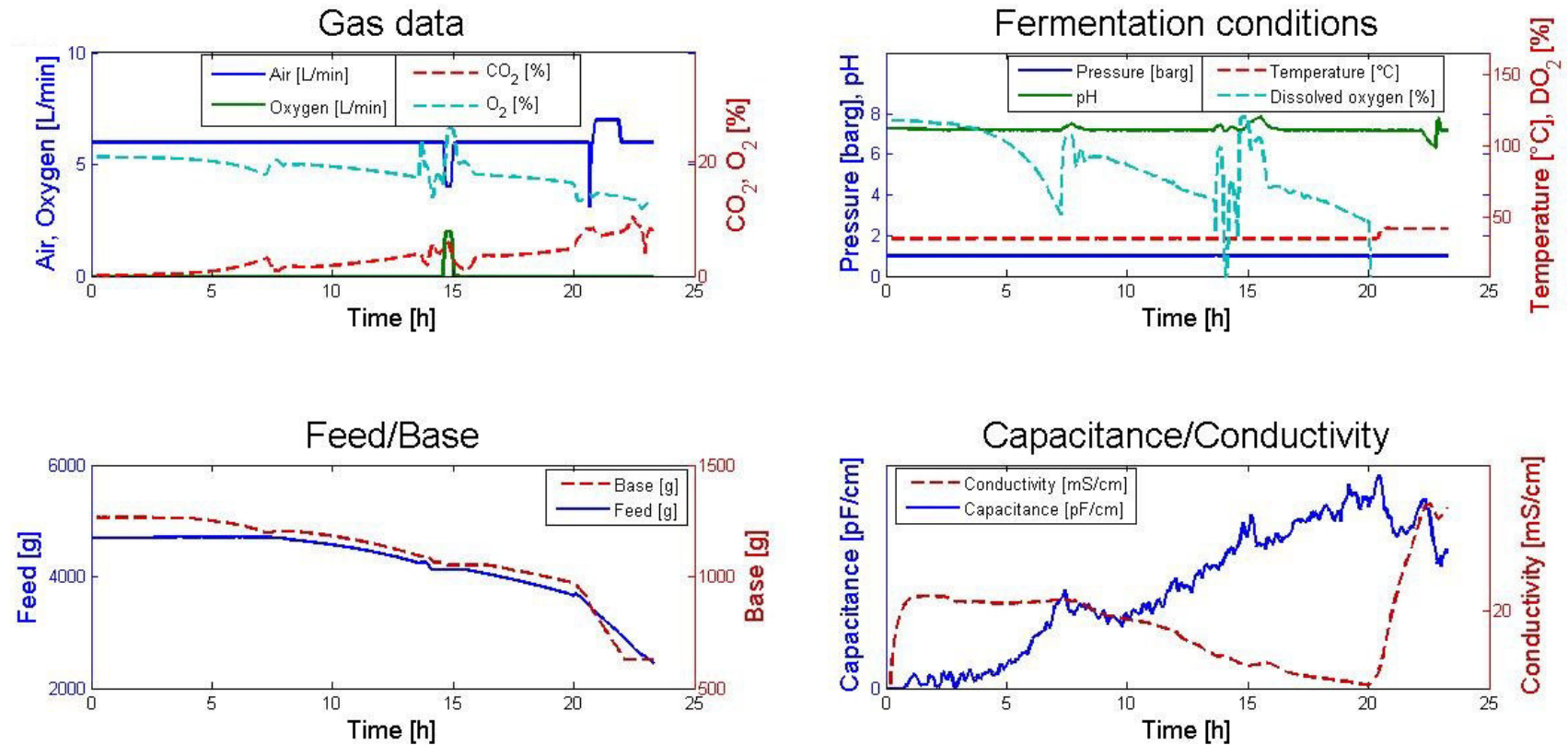


Figure 71: Online data of fermentation 7. Gas data like air inflow (blue), oxygen inflow (green), CO₂ (red) and O₂ (cyan) in the off-gas are shown in the graph on the top left. On the top right fermenter pressure (blue), pH (green), fermentation temperature (red) and dissolved oxygen (cyan) are seen. In the bottom graphs feed (blue) and base (red) balance weights on the left are presented as well as capacitance (blue) and conductivity (red) measurements on the right.

The oxygen supply in fermentation 7 was again controlled by an external mass flow controller. Therefore the oxygen signal in Figure 71 is not representative.

Offline data of fermentation 7:

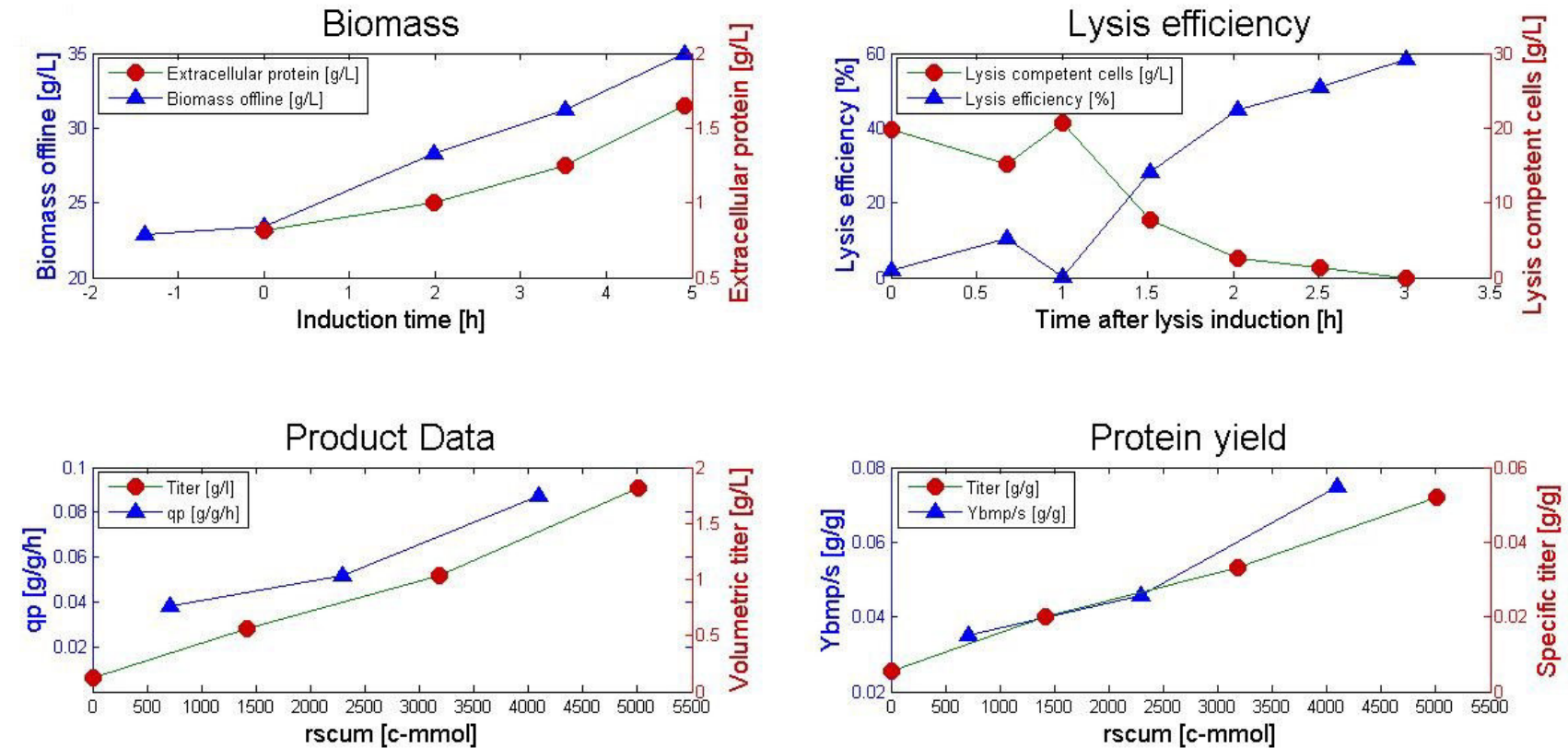


Figure 72: Offline data of fermentation 7. On the top left biomass concentration (blue) and extracellular protein concentration (red) are shown over induction time. The beginning of mixed feed was defined as starting point. Lysis efficiency (blue) and lysis competent cells (red) are plotted in the graph on the top right. The graphs on the bottom present specific rhBMP-2 production rate (blue) and the volumetric rhBMP-2 titer (red) on the left as well as the rhBMP-2 yield (blue) and the specific rhBMP-2 titer (red) on the right. On the abscissa in the two last mentioned graphs, there is cumulated substrate. With one exception offline data is only shown for induction phase. Lysis efficiency and lysis competent cells are plotted for lysis phase.

Processed data of fermentation 7:

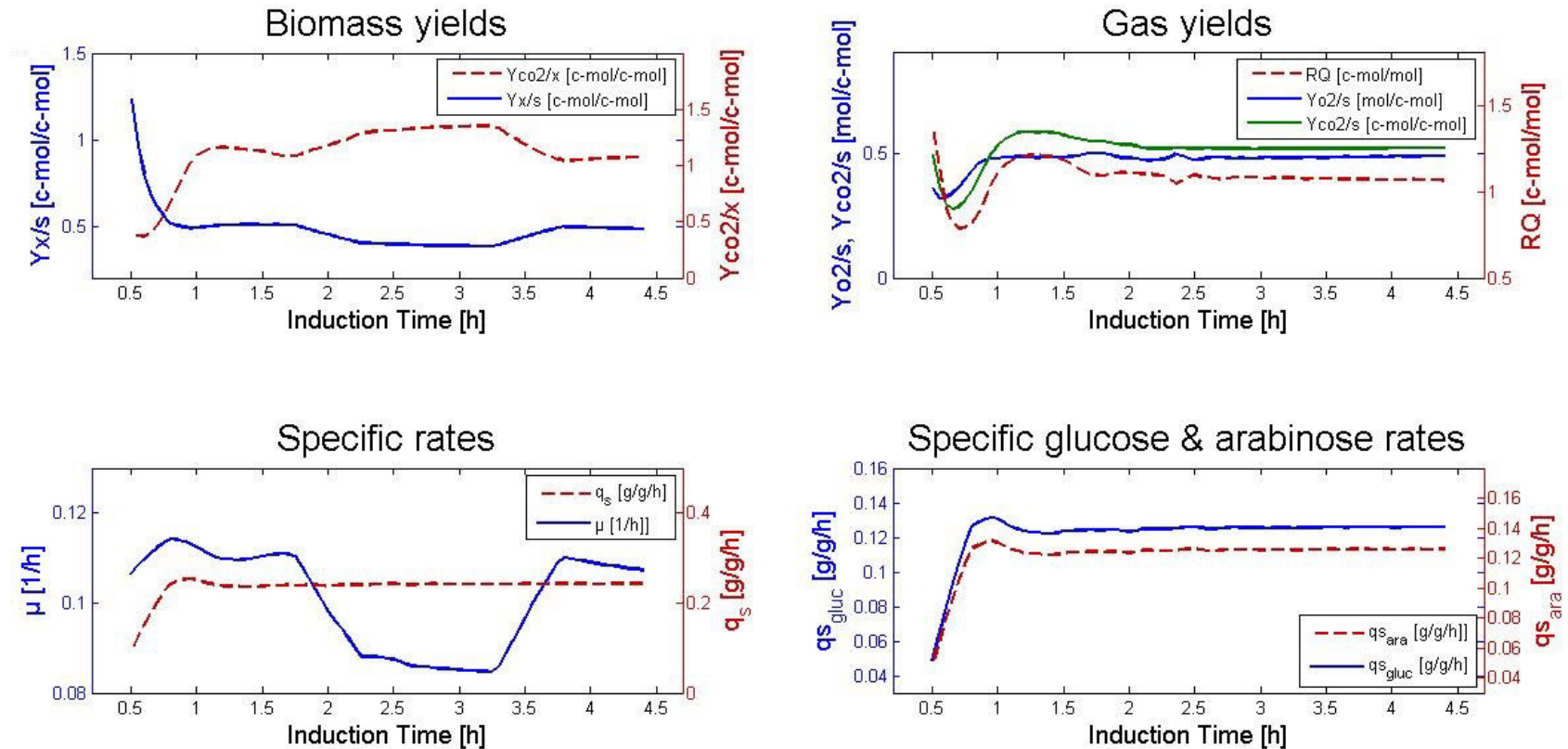


Figure 73: Processed data of fermentation 7. Biomass yield (blue) and $Y_{CO_2/s}$ (red) are plotted in the graph on the top left. Next to it the gas yields are shown ($Y_{O_2/s}$ in blue, $Y_{CO_2/s}$ in green and the respiratory quotient in red). On the bottom the total specific substrate uptake rate (red) and the specific growth rate (blue) during recombinant protein production are found in the left, whereas specific D-glucose uptake rate (blue) and specific L-arabinose uptake rate (red) are on the bottom right. Processed data is only shown for the induction phase.

Online data of fermentation 8:

Process parameters: $q_{s_{gluc}}=0.125\text{g/g/h}$, $q_{s_{ara}}=0.2\text{g/g/h}$ (set-points)

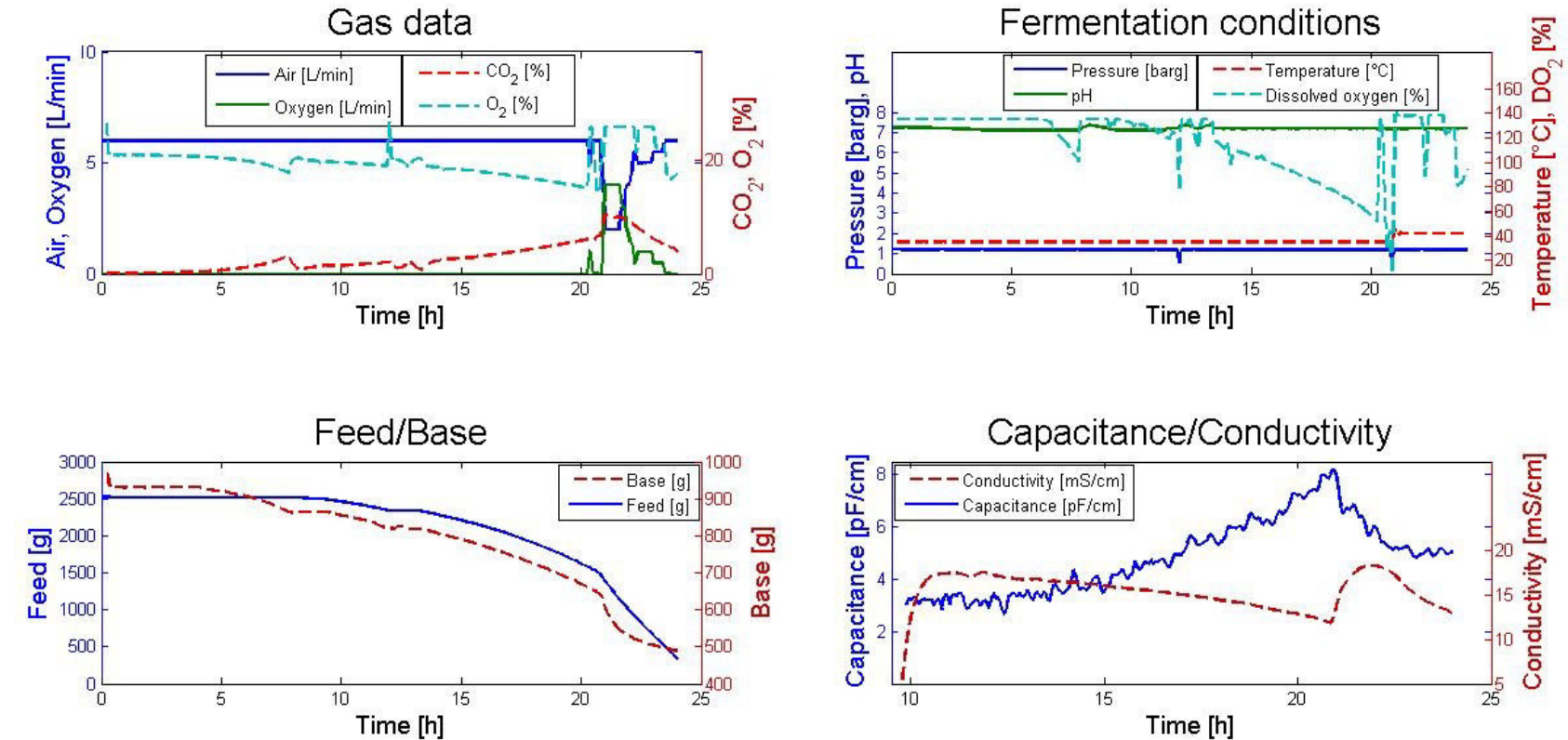


Figure 74: Online data of fermentation 8. Gas data like air inflow (blue), oxygen inflow (green), CO_2 (red) and O_2 (cyan) in the off-gas are shown in the graph on the top left. On the top right fermenter pressure (blue), pH (green), fermentation temperature (red) and dissolved oxygen (cyan) are seen. In the bottom graphs feed (blue) and base (red) balance weights on the left are presented as well as capacitance (blue) and conductivity (red) measurements on the right.

The first 10 hours of fermentation, the capacitance probe was not appropriately connected. Therefore capacitance and conductivity measurements were started 10 hours after inoculation (see Figure 74).

Offline data of fermentation 8:

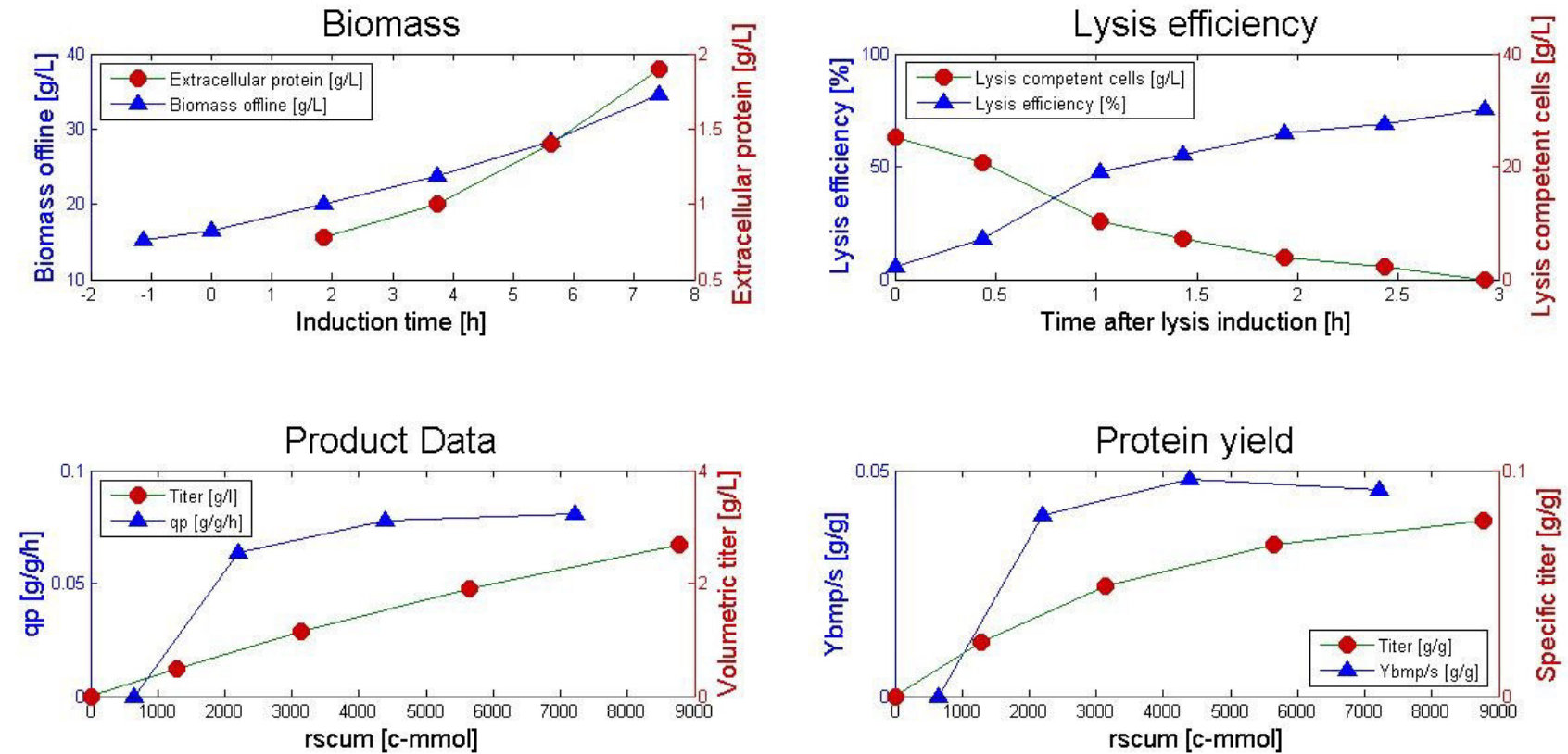


Figure 75: Offline data of fermentation 8. On the top left biomass concentration (blue) and extracellular protein concentration (red) are shown over induction time. The beginning of mixed feed was defined as starting point. Lysis efficiency (blue) and lysis competent cells (red) are plotted in the graph on the top right. The graphs on the bottom present specific rhBMP-2 production rate (blue) and the volumetric rhBMP-2 titer (red) on the left as well as the rhBMP-2 yield (blue) and the specific rhBMP-2 titer (red) on the right. On the abscissa in the two last mentioned graphs, there is cumulated substrate. With one exception offline data is only shown for induction phase. Lysis efficiency and lysis competent cells are plotted for lysis phase.

Processed data of fermentation 8:

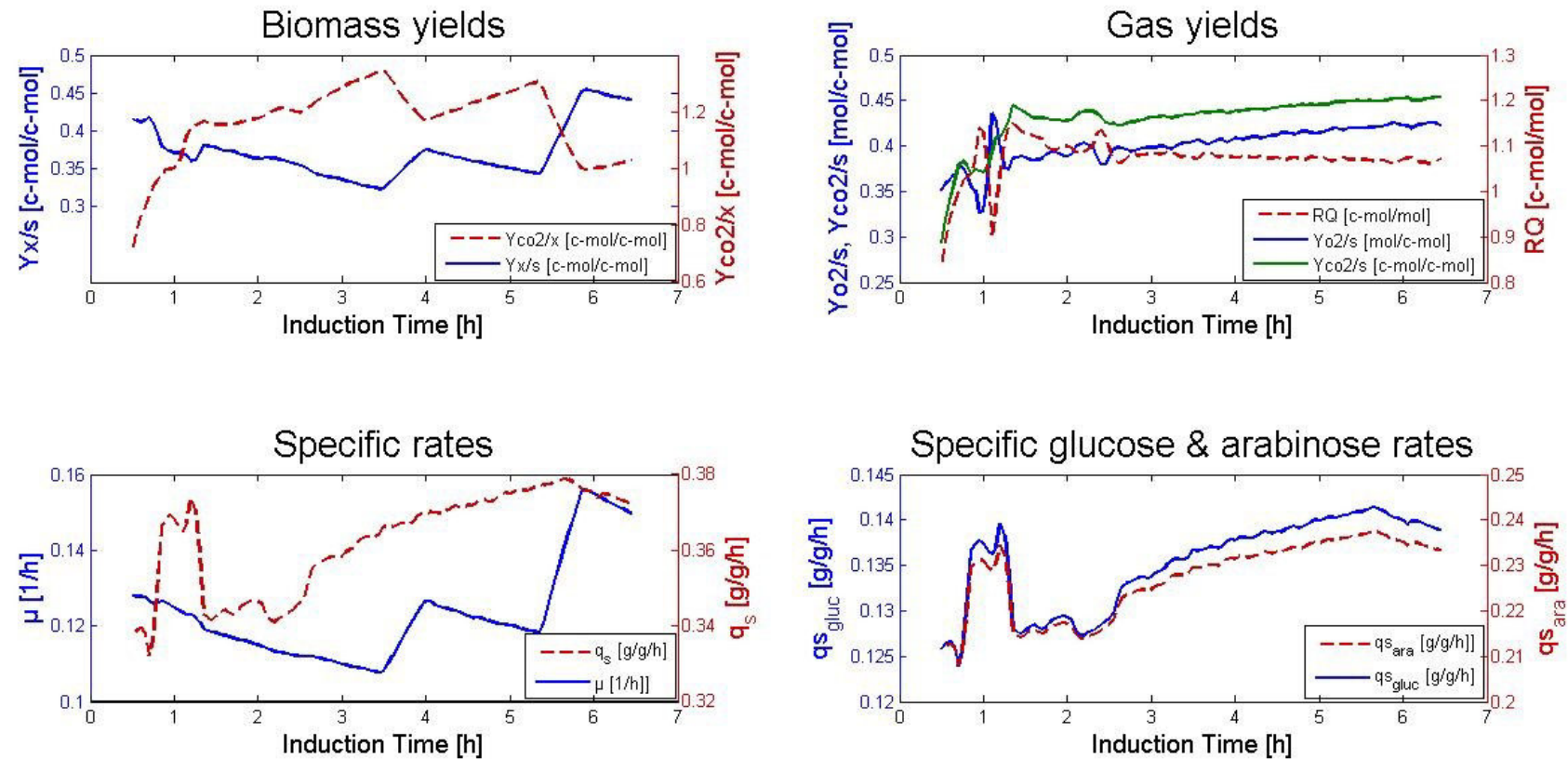


Figure 76: Processed data of fermentation 8. Biomass yield (blue) and $Y_{CO_2/s}$ (red) are plotted in the graph on the top left. Next to it the gas yields are shown ($Y_{O_2/s}$ in blue, $Y_{CO_2/s}$ in green and the respiratory quotient in red). On the bottom the total specific substrate uptake rate (red) and the specific growth rate (blue) during recombinant protein production are found in the left, whereas specific D-glucose uptake rate (blue) and specific L-arabinose uptake rate (red) are on the bottom right. Processed data is only shown for the induction phase.

Online data of fermentation 9:

Process parameters: $q_{s_{gluc}}=0.075\text{g/g/h}$, $q_{s_{ara}}=0.075\text{g/g/h}$ (set-points)

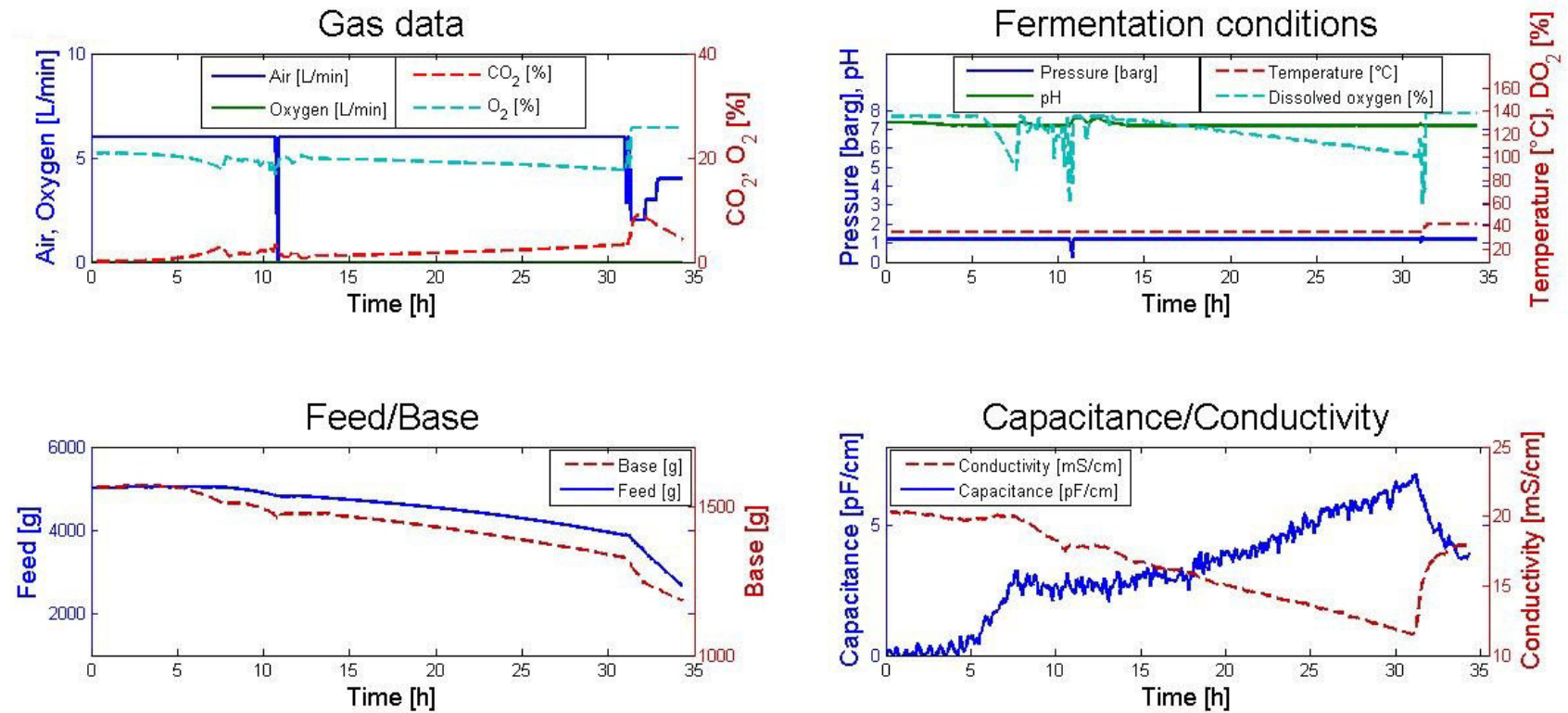


Figure 77: Online data of fermentation 9. Gas data like air inflow (blue), oxygen inflow (green), CO_2 (red) and O_2 (cyan) in the off-gas are shown in the graph on the top left. On the top right fermenter pressure (blue), pH (green), fermentation temperature (red) and dissolved oxygen (cyan) are seen. In the bottom graphs feed (blue) and base (red) balance weights on the left are presented as well as capacitance (blue) and conductivity (red) measurements on the right.

Offline data of fermentation 9:

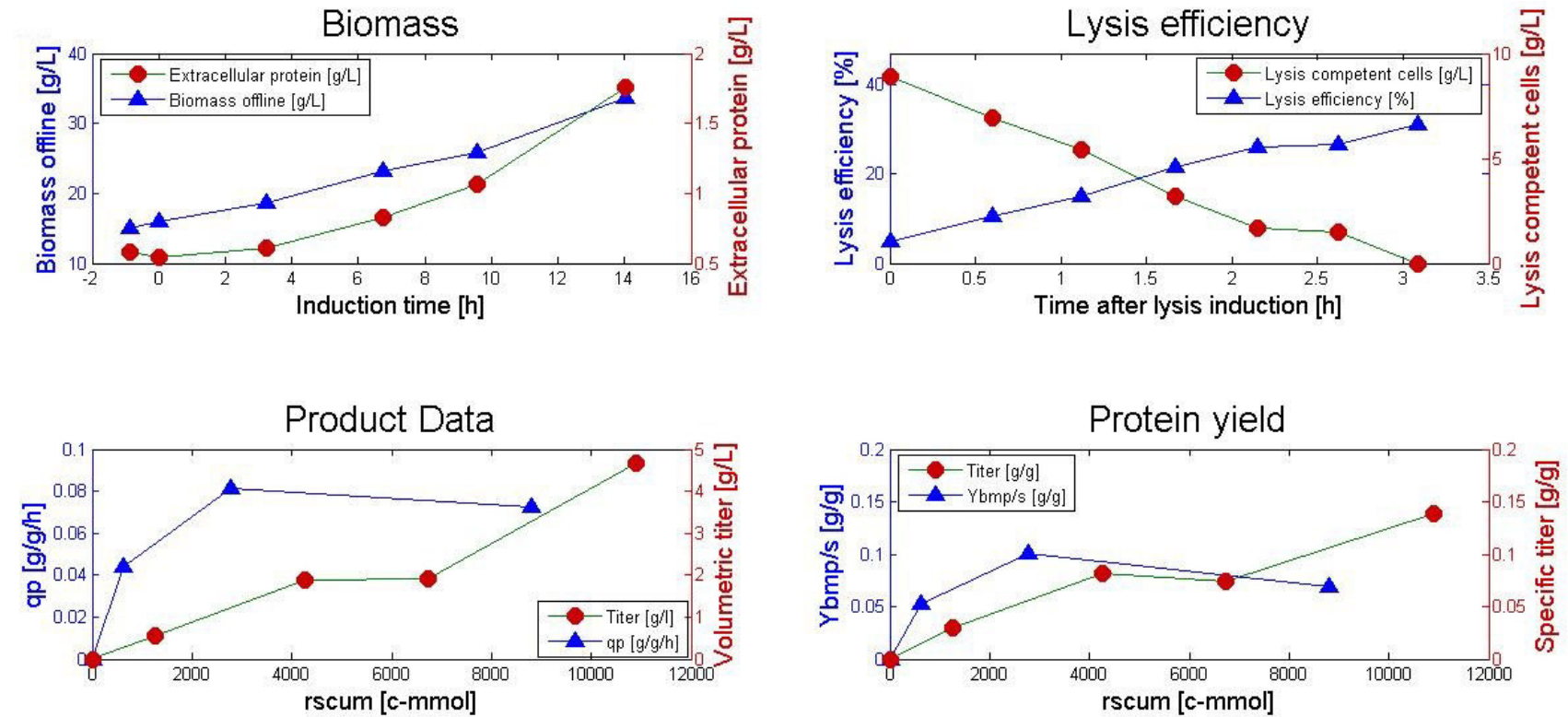


Figure 78: Offline data of fermentation 9. On the top left biomass concentration (blue) and extracellular protein concentration (red) are shown over induction time. The beginning of mixed feed was defined as starting point. Lysis efficiency (blue) and lysis competent cells (red) are plotted in the graph on the top right. The graphs on the bottom present specific rhBMP-2 production rate (blue) and the volumetric rhBMP-2 titer (red) on the left as well as the rhBMP-2 yield (blue) and the specific rhBMP-2 titer (red) on the right. On the abscissa in the two last mentioned graphs, there is cumulated substrate. With one exception offline data is only shown for induction phase. Lysis efficiency and lysis competent cells are plotted for lysis phase.

Processed data of fermentation 9:

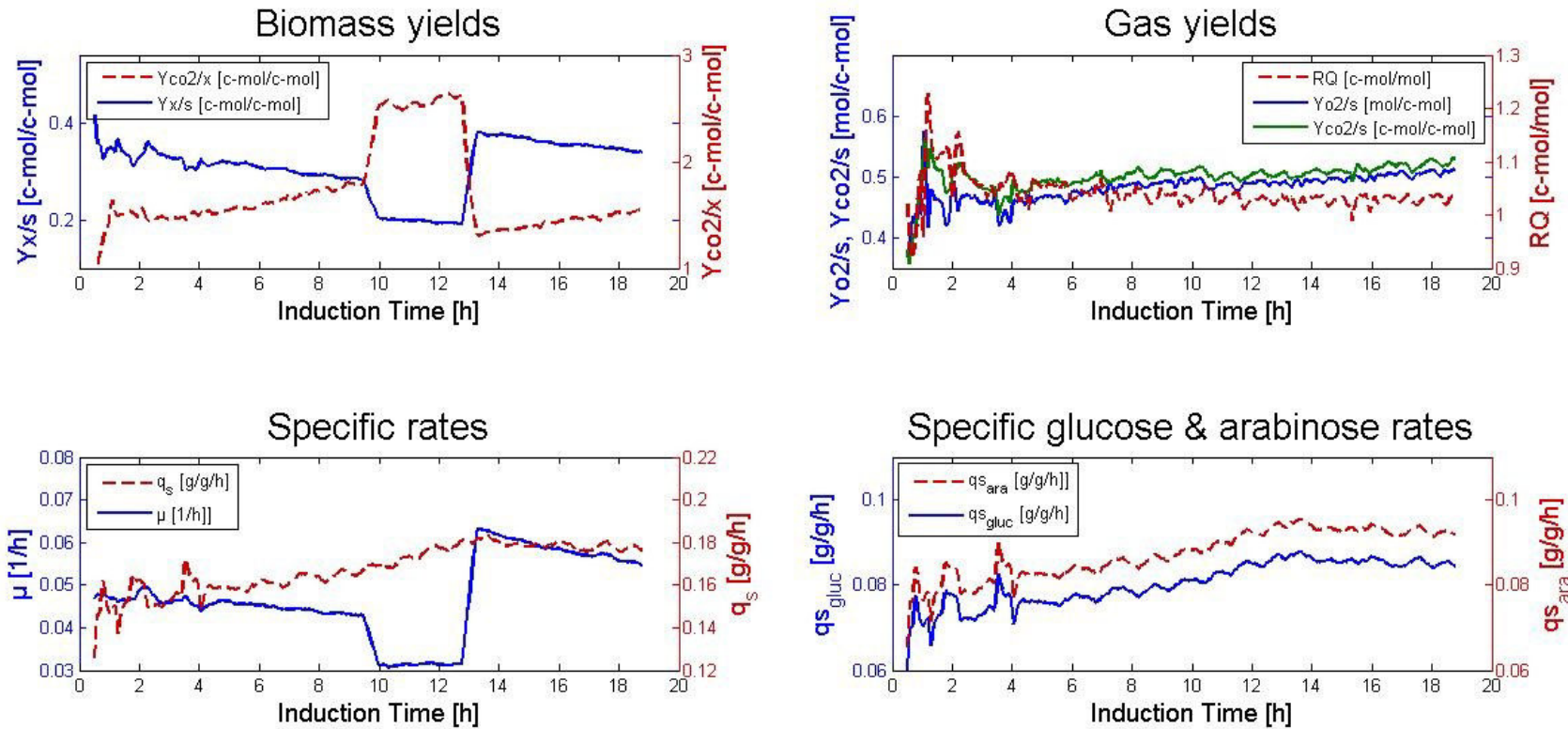


Figure 79: Processed data of fermentation 9. Biomass yield (blue) and $Y_{CO_2/s}$ (red) are plotted in the graph on the top left. Next to it the gas yields are shown ($Y_{O_2/s}$ in blue, $Y_{CO_2/s}$ in green and the respiratory quotient in red). On the bottom the total specific substrate uptake rate (red) and the specific growth rate (blue) during recombinant protein production are found in the left, whereas specific D-glucose uptake rate (blue) and specific L-arabinose uptake rate (red) are on the bottom right. Processed data is only shown for the induction phase.

Online data of fermentation 10:

Process parameters: $q_{s_{gluc}}=0.05\text{g/g/h}$, $q_{s_{ara}}=0.125\text{g/g/h}$

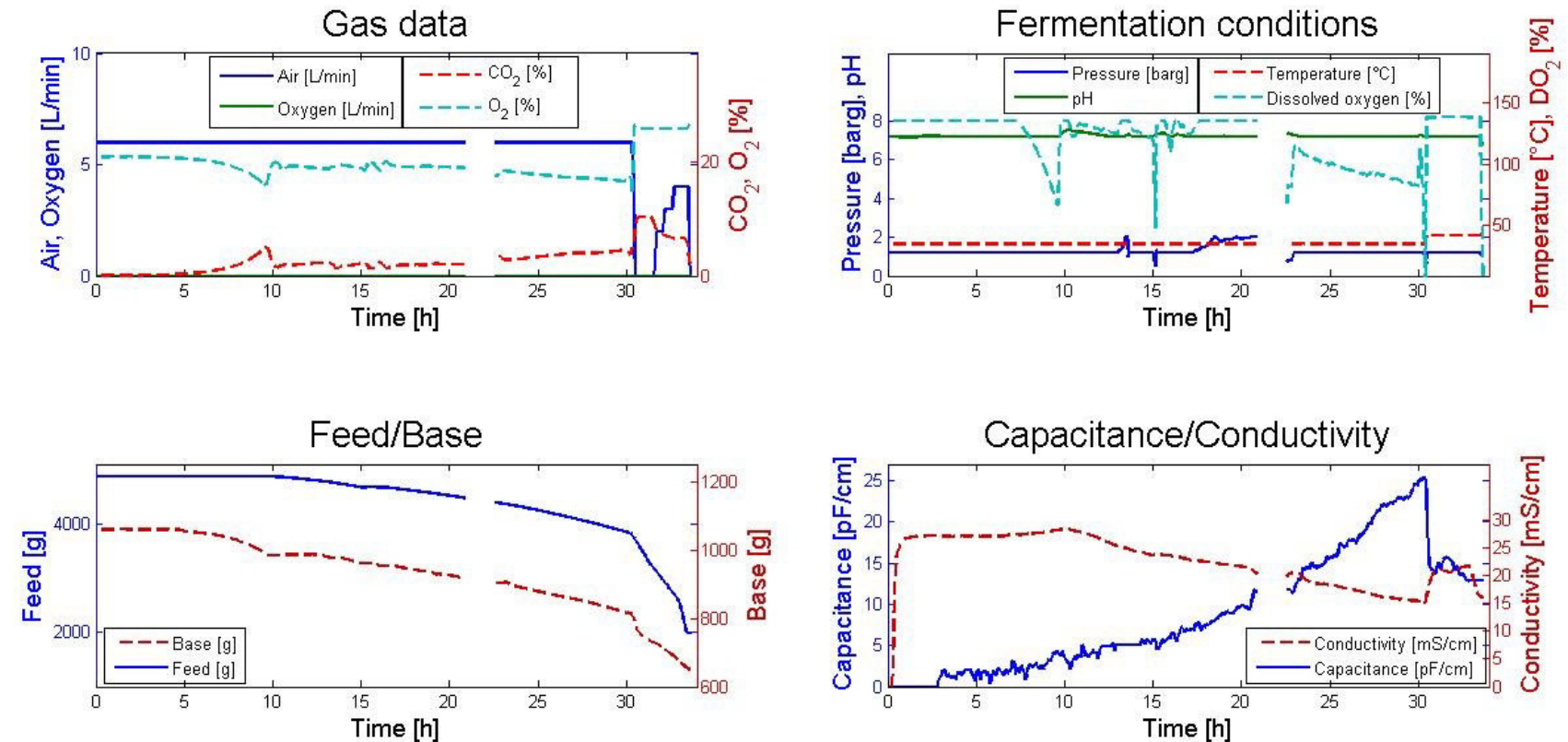


Figure 80: Online data of fermentation 10. Gas data like air inflow (blue), oxygen inflow (green), CO_2 (red) and O_2 (cyan) in the off-gas are shown in the graph on the top left. On the top right fermenter pressure (blue), pH (green), fermentation temperature (red) and dissolved oxygen (cyan) are seen. In the bottom graphs feed (blue) and base (red) balance weights on the left are presented as well as capacitance (blue) and conductivity (red) measurements on the right.

In fermentation 10 the process control system stopped for one and a half hours. In this time process parameters neither could be changed nor were the signals recorded. This stop of recording is seen in Figures 80 and 82.

Offline data of fermentation 10:

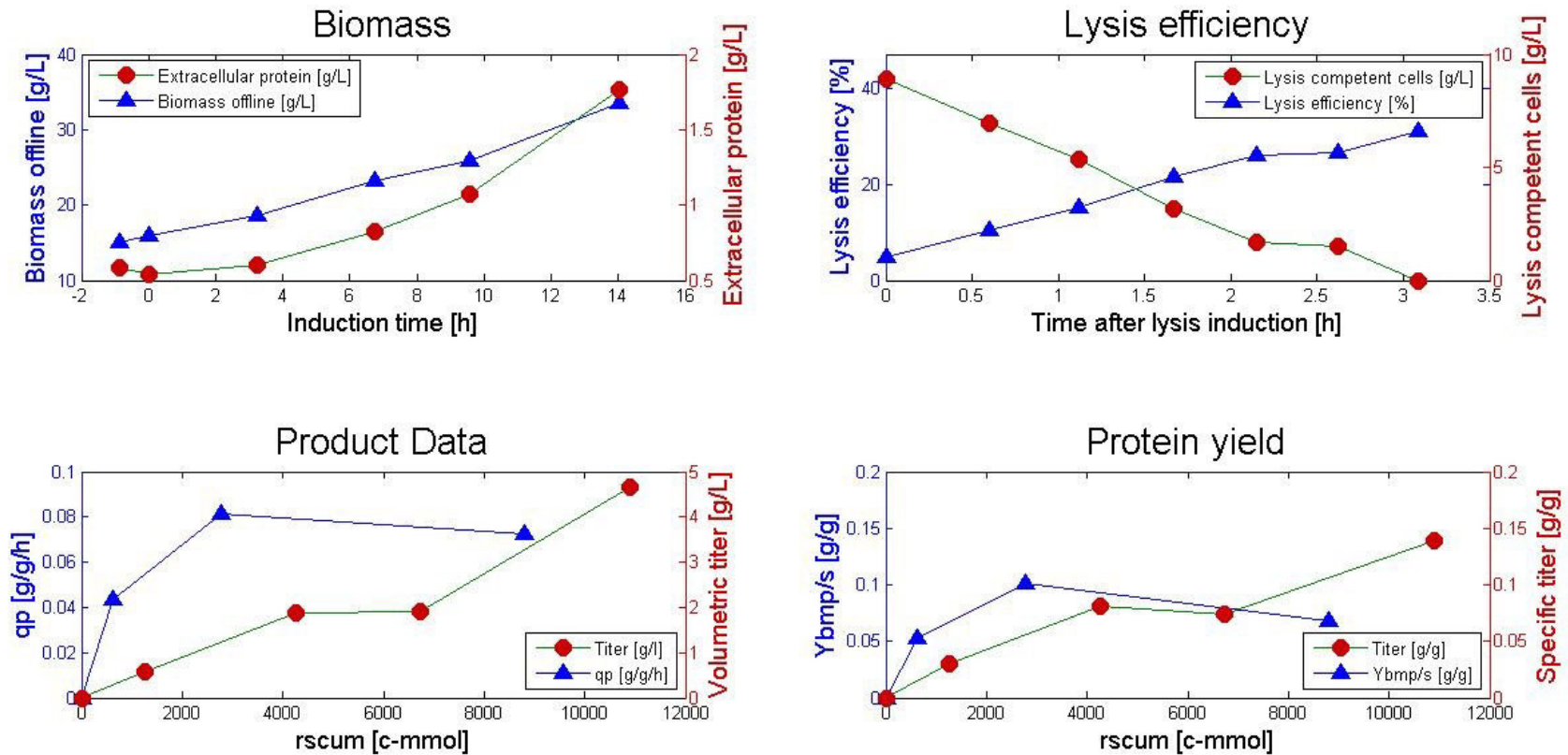


Figure 81: Offline data of fermentation 10. On the top left biomass concentration (blue) and extracellular protein concentration (red) are shown over induction time. The beginning of mixed feed was defined as starting point. Lysis efficiency (blue) and lysis competent cells (red) are plotted in the graph on the top right. The graphs on the bottom present specific rhBMP-2 production rate (blue) and the volumetric rhBMP-2 titer (red) on the left as well as the rhBMP-2 yield (blue) and the specific rhBMP-2 titer (red) on the right. On the abscissa in the two last mentioned graphs, there is cumulated substrate. With one exception offline data is only shown for induction phase. Lysis efficiency and lysis competent cells are plotted for lysis phase.

Processed data of fermentation 10:

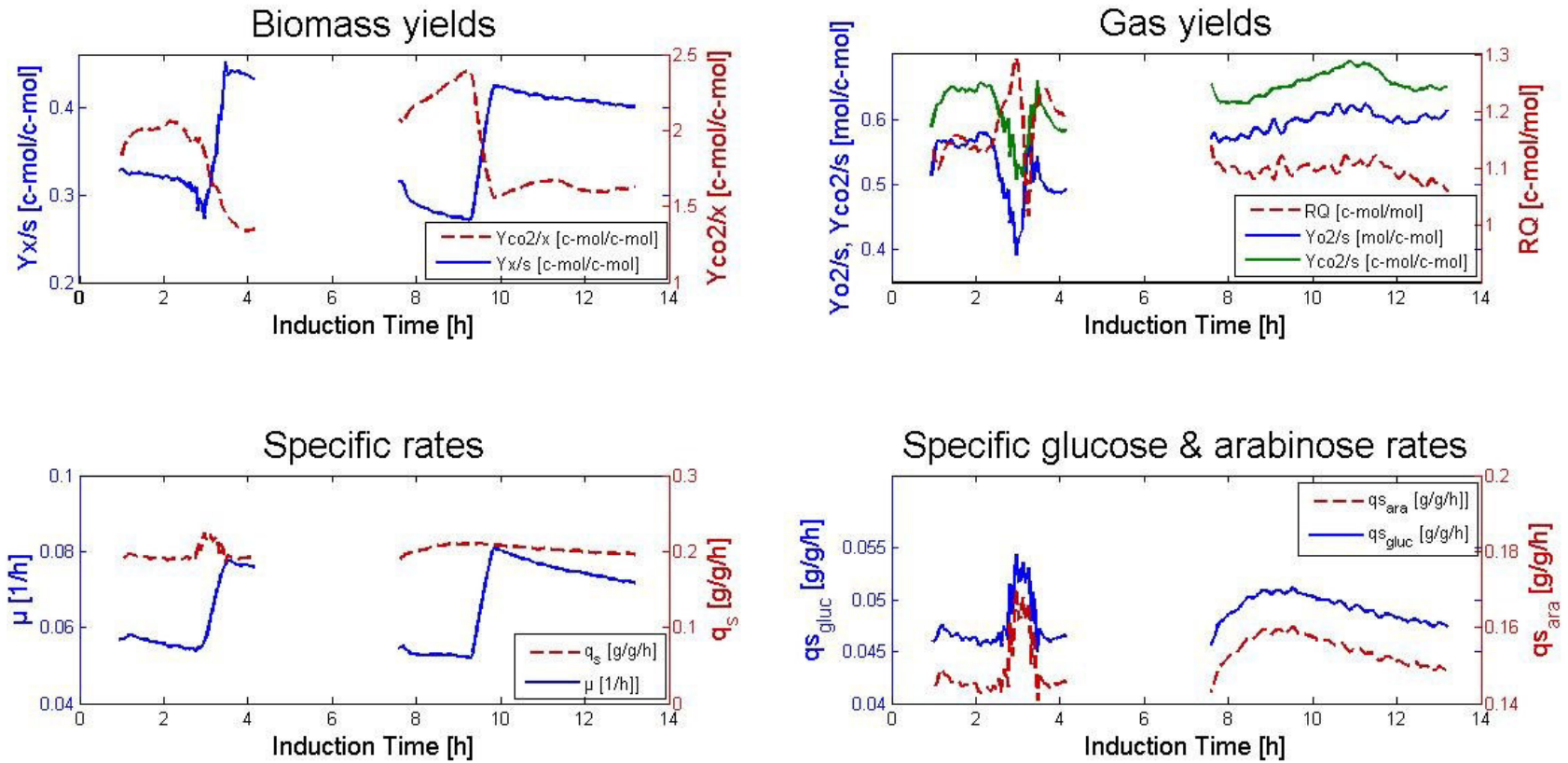


Figure 82: Processed data of fermentation 10. Biomass yield (blue) and $Y_{CO_2/s}$ (red) are plotted in the graph on the top left. Next to it the gas yields are shown ($Y_{O_2/s}$ in blue, $Y_{CO_2/s}$ in green and the respiratory quotient in red). On the bottom the total specific substrate uptake rate (red) and the specific growth rate (blue) during recombinant protein production are found in the left, whereas specific D-glucose uptake rate (blue) and specific L-arabinose uptake rate (red) are on the bottom right. Processed data is only shown for the induction phase.

7. References

1. Demain AL, Vaishnav P: **Production of recombinant proteins by microbes and higher organisms.** *Biotechnology advances* 2009, **27**(3):297-306.
2. Berkowitz SA, Engen JR, Mazzeo JR, Jones GB: **Analytical tools for characterizing biopharmaceuticals and the implications for biosimilars.** *Nature reviews Drug discovery* 2012, **11**(7):527-540.
3. Jha S, Agarwal S, Sanyal I, Jain GK, Amla DV: **Differential subcellular targeting of recombinant human alpha(1)-proteinase inhibitor influences yield, biological activity and in planta stability of the protein in transgenic tomato plants.** *Plant science : an international journal of experimental plant biology* 2012, **196**:53-66.
4. Houdebine LM: **Production of pharmaceutical proteins by transgenic animals.** *Comparative immunology, microbiology and infectious diseases* 2009, **32**(2):107-121.
5. Ren X, Yu D, Yu L, Gao G, Han S, Feng Y: **A new study of cell disruption to release recombinant thermostable enzyme from Escherichia coli by thermolysis.** *Journal of biotechnology* 2007, **129**(4):668-673.
6. Lubitz W, Witte A, Eko FO, Kamal M, Jechlinger W, Brand E, Marchart J, Haidinger W, Huter V, Felnerova D *et al*: **Extended recombinant bacterial ghost system.** *Journal of biotechnology* 1999, **73**(2-3):261-273.
7. Langemann T, Koller VJ, Muhammad A, Kudela P, Mayr UB, Lubitz W: **The Bacterial Ghost platform system: production and applications.** *Bioengineered bugs* 2010, **1**(5):326-336.
8. Sagmeister P, Schimek C, Meitz A, Herwig C, Spadiut O: **Design and multivariate investigation of an E. coli pBAD mixed fed fed-batch expression platform capable of tunable recombinant protein expression.** Submitted manuscript.
9. Andersen DC, Krummen L: **Recombinant protein expression for therapeutic applications.** *Current opinion in biotechnology* 2002, **13**(2):117-123.
10. Carneiro S, Ferreira EC, Rocha I: **Metabolic responses to recombinant bioprocesses in Escherichia coli.** *Journal of biotechnology* 2012.
11. Choi JH, Keumb KC, Lee S: **Production of recombinant proteins by high cell density culture of Escherichia coli.** *Chemical Engineering Science* 2006, **61**:876 – 885.
12. Swartz JR: **Advances in Escherichia coli production of therapeutic proteins.** *Current opinion in biotechnology* 2001, **12**(2):195-201.
13. Baneyx F, Mujacic M: **Recombinant protein folding and misfolding in Escherichia coli.** *Nature biotechnology* 2004, **22**(11):1399-1408.
14. Sorensen HP, Mortensen KK: **Advanced genetic strategies for recombinant protein expression in Escherichia coli.** *Journal of biotechnology* 2005, **115**(2):113-128.
15. Jana S, Deb JK: **Strategies for efficient production of heterologous proteins in Escherichia coli.** *Applied microbiology and biotechnology* 2005, **67**:289–298.

16. Miroux B, Walker JE: **Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels.** *Journal of molecular biology* 1996, **260**(3):289-298.
17. Sagmeister P, Wechselberger P, Jazini M, Meitz A, Langemann T, Herwig C: **Soft sensor assisted dynamic bioprocess control: Efficient tools for bioprocess development.** *Chemical Engineering Science* 2013, **96**:190-198.
18. Bird PI, Pak SC, Worrall DM, Bottomley SP: **Production of recombinant serpins in Escherichia coli.** *Methods* 2004, **32**:169–176.
19. Sanden AM, Prytz I, Tubulekas I, Forberg C, Le H, Hektor A, Neubauer P, Pragai Z, Harwood C, Ward A *et al*: **Limiting factors in Escherichia coli fed-batch production of recombinant proteins.** *Biotechnology and bioengineering* 2003, **81**(2):158-166.
20. Ami D, Natalello A, Gatti-Lafranconi P, Lotti M, Doglia SM: **Kinetics of inclusion body formation studied in intact cells by FT-IR spectroscopy.** *FEBS letters* 2005, **579**(16):3433-3436.
21. Margreiter G, Messner P, Caldwell KD, Bayer K: **Size characterization of inclusion bodies by sedimentation field-flow fractionation.** *Journal of biotechnology* 2008, **138**(3-4):67-73.
22. Choi JH, Lee SY: **Secretory and extracellular production of recombinant proteins using Escherichia coli.** *Applied microbiology and biotechnology* 2004, **64**(5):625-635.
23. Bäcklund E: **Impact of glucose uptake rate on recombinant protein production in Escherichia coli.** Stockholm; 2011.
24. Anjou MC, Daugulis AJ: **Mixed-feed exponential feeding for fed-batch culture of recombinant methylotrophic yeast.** *Biotechnology Letters* 2000, **22**:341–346.
25. Eiteman MA, Altman E: **Overcoming acetate in Escherichia coli recombinant protein fermentations.** *Trends in biotechnology* 2006, **24**(11):530-536.
26. Phue JN, Shiloach J: **Impact of dissolved oxygen concentration on acetate accumulation and physiology of E. coli BL21, evaluating transcription levels of key genes at different dissolved oxygen conditions.** *Metabolic engineering* 2005, **7**(5-6):353-363.
27. Paczia N, Nilgen A, Lehmann T, Gatgens J, Wiechert W, Noack S: **Extensive exometabolome analysis reveals extended overflow metabolism in various microorganisms.** *Microbial cell factories* 2012, **11**:122.
28. Papagianni M: **Recent advances in engineering the central carbon metabolism of industrially important bacteria.** *Microbial cell factories* 2012, **11**:50.
29. Heyland J, Blank LM, Schmid A: **Quantification of metabolic limitations during recombinant protein production in Escherichia coli.** *Journal of biotechnology* 2011, **155**(2):178-184.
30. Bostrom M, Markland K, Sanden AM, Hedhammar M, Hober S, Larsson G: **Effect of substrate feed rate on recombinant protein secretion, degradation and inclusion body formation in Escherichia coli.** *Applied microbiology and biotechnology* 2005, **68**(1):82-90.
31. Ow DS-W, Nissom PM, Philp R, Oha SK-W, Yap MG-S: **Global transcriptional analysis of metabolic burden due to plasmid maintenance in Escherichia coli DH5 during batch fermentation.** *Enzyme and Microbial Technology* 2006, **39**:391–398.

32. Yang M, Johnson SC, Murthy PP: **Enhancement of alkaline phytase production in *Pichia pastoris*: influence of gene dosage, sequence optimization and expression temperature.** *Protein expression and purification* 2012, **84**(2):247-254.
33. Ami D, Natalello A, Schultz T, Gatti-Lafranconi P, Lotti M, Doglia SM, de Marco A: **Effects of recombinant protein misfolding and aggregation on bacterial membranes.** *Biochimica et biophysica acta* 2009, **1794**(2):263-269.
34. Wozney JM: **Overview of bone morphogenetic proteins.** *Spine (Phila Pa 1976)* 2002, **27**(16 Suppl 1):S2-8.
35. FDA: **Guidance for Industry: PAT - a Framework for Innovative Pharmaceutical Manufacturing and Quality Assurance.** In. Rockville: Food and Drug Administration; 2004.
36. HARMONISATION ICO: **PHARMACEUTICAL DEVELOPMENT Q8 (R2).** In.; 2009.
37. FDA: **Guidance for Industry: Q9 Quality Risk Management.** In. Rockville: Food and Drug Administration; 2006.
38. HARMONISATION ICO: **PHARMACEUTICAL DEVELOPMENT: ANNEX TO Q8.** In.; 2007.
39. Chen D, Zhao M, Mundy GR: **Bone morphogenetic proteins.** *Growth Factors* 2004, **22**(4):233-241.
40. **BMP 2 Human** [http://www.prospecbio.com/BMP-2_Human_3_4/]
41. Rhee JI, Bode J, Diaz-Ricci JC, Poock D, Weigel B, Kretzmer G, Schugler K: **Influence of the medium composition and plasmid combination on the growth of recombinant *Escherichia coli* JM109 and on the production of the fusion protein EcoRI::SPA.** *Journal of biotechnology* 1997, **55**(2):69-83.
42. Zhang Y, Ma Y, Yang M, Min S, Yao J, Zhu L: **Expression, purification, and refolding of a recombinant human bone morphogenetic protein 2 in vitro.** *Protein expression and purification* 2011, **75**(2):155-160.
43. Schleif R: **Regulation of the L-arabinose operon of *Escherichia coli*.** *Trends in genetics : TIG* 2000, **16**(12):559-565.
44. Englesberg E, Irr J, Power J, Lee N: **Positive control of enzyme synthesis by gene C in the L-arabinose system.** *Journal of bacteriology* 1965, **90**(4):946-957.
45. Englesberg E, Squires C, Meronk F, Jr.: **The L-arabinose operon in *Escherichia coli* B-r: a genetic demonstration of two functional states of the product of a regulator gene.** *Proceedings of the National Academy of Sciences of the United States of America* 1969, **62**(4):1100-1107.
46. Johnson CM, Schleif RF: **In vivo induction kinetics of the arabinose promoters in *Escherichia coli*.** *Journal of bacteriology* 1995, **177**(12):3438-3442.
47. Schleif R: **AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action.** *FEMS microbiology reviews* 2010, **34**(5):779-796.
48. Hirsh J, Schleif R: **The araC promoter: transcription, mapping and interaction with the araBAD promoter.** *Cell* 1977, **11**(3):545-550.
49. Horazdovsky BF, Hogg RW: **Genetic reconstitution of the high-affinity L-arabinose transport system.** *Journal of bacteriology* 1989, **171**(6):3053-3059.
50. Lobell RB, Schleif RF: **DNA looping and unlooping by AraC protein.** *Science* 1990, **250**(4980):528-532.

51. Lee DH, Huo L, Schleif R: **Repression of the araBAD promoter from araO1.** *Journal of molecular biology* 1992, **224**(2):335-341.
52. Schleif R: **AraC protein: a love-hate relationship.** *BioEssays : news and reviews in molecular, cellular and developmental biology* 2003, **25**(3):274-282.
53. Guzman LM, Belin D, Carson MJ, Beckwith J: **Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter.** *Journal of bacteriology* 1995, **177**(14):4121-4130.
54. Siegele DA, Hu JC: **Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations.** *Proceedings of the National Academy of Sciences of the United States of America* 1997, **94**(15):8168-8172.
55. Jungo C, Marison I, von Stockar U: **Mixed feeds of glycerol and methanol can improve the performance of Pichia pastoris cultures: A quantitative study based on concentration gradients in transient continuous cultures.** *Journal of biotechnology* 2007, **128**(4):824-837.
56. Striedner G, Cserjan-Puschmann M, Potschacher F, Bayer K: **Tuning the transcription rate of recombinant protein in strong Escherichia coli expression systems through repressor titration.** *Biotechnology progress* 2003, **19**(5):1427-1432.
57. Sagmeister P, Kment M, Wechselberger P, Meitz A, Langemann T, Herwig C: **Soft-sensor assisted dynamic investigation of mixed feed bioprocesses.** In.; Submitted manuscript.
58. Khlebnikov A, Datsenko KA, Skaug T, Wanner BL, Keasling JD: **Homogeneous expression of the P(BAD) promoter in Escherichia coli by constitutive expression of the low-affinity high-capacity AraE transporter.** *Microbiology* 2001, **147**(Pt 12):3241-3247.
59. Khlebnikov A, Risa O, Skaug T, Carrier TA, Keasling JD: **Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture.** *Journal of bacteriology* 2000, **182**(24):7029-7034.
60. Khlebnikov A, Skaug T, Keasling JD: **Modulation of gene expression from the arabinose-inducible araBAD promoter.** *Journal of industrial microbiology & biotechnology* 2002, **29**(1):34-37.
61. Morgan-Kiss RM, Wadler C, Cronan JE, Jr.: **Long-term and homogeneous regulation of the Escherichia coli araBAD promoter by use of a lactose transporter of relaxed specificity.** *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**(11):7373-7377.
62. Jungo C, Schenk J, Pasquier M, Marison IW, von Stockar U: **A quantitative analysis of the benefits of mixed feeds of sorbitol and methanol for the production of recombinant avidin with Pichia pastoris.** *Journal of biotechnology* 2007, **131**(1):57-66.
63. Zalai D, Dietzsch C, Herwig C, Spadiut O: **A dynamic fed batch strategy for a Pichia pastoris mixed feed system to increase process understanding.** *Biotechnology progress* 2012, **28**(3):878-886.
64. Haslberger AG, Kohl G, Felnerova D, Mayr UB, Furst-Ladani S, Lubitz W: **Activation, stimulation and uptake of bacterial ghosts in antigen presenting cells.** *Journal of biotechnology* 2000, **83**(1-2):57-66.
65. Witte A, Blasi U, Halfmann G, Szostak M, Wanner G, Lubitz W: **Phi X174 protein E-mediated lysis of Escherichia coli.** *Biochimie* 1990, **72**(2-3):191-200.

66. Witte A, Wanner G, Blasi U, Halfmann G, Szostak M, Lubitz W: **Endogenous transmembrane tunnel formation mediated by phi X174 lysis protein E.** *Journal of bacteriology* 1990, **172**(7):4109-4114.
67. Schon P, Schrot G, Wanner G, Lubitz W, Witte A: **Two-stage model for integration of the lysis protein E of phi X174 into the cell envelope of Escherichia coli.** *FEMS microbiology reviews* 1995, **17**(1-2):207-212.
68. Szostak MP, Hensel A, Eko FO, Klein R, Auer T, Mader H, Haslberger A, Bunka S, Wanner G, Lubitz W: **Bacterial ghosts: non-living candidate vaccines.** *Journal of biotechnology* 1996, **44**(1-3):161-170.
69. Sagmeister P, Langemann T, Herwig C: **Final Report WP 02: Development of a production process for Bacterial Ghosts along QbD principles (USP & DSP).** In. Vienna: RCPE; 2012.
70. Hoeffelner H, Haas R: **Recombinant bacterial ghosts: versatile targeting vehicles and promising vaccine candidates.** *International journal of medical microbiology : IJMM* 2004, **294**(5):303-311.
71. Tabrizi CA, Walcher P, Mayr UB, Stiedl T, Binder M, McGrath J, Lubitz W: **Bacterial ghosts--biological particles as delivery systems for antigens, nucleic acids and drugs.** *Current opinion in biotechnology* 2004, **15**(6):530-537.
72. Kudela P, Koller VJ, Lubitz W: **Bacterial ghosts (BGs)--advanced antigen and drug delivery system.** *Vaccine* 2010, **28**(36):5760-5767.
73. Ebensen T, Paukner S, Link C, Kudela P, de Domenico C, Lubitz W, Guzman CA: **Bacterial ghosts are an efficient delivery system for DNA vaccines.** *J Immunol* 2004, **172**(11):6858-6865.
74. Paukner S, Kohl G, Lubitz W: **Bacterial ghosts as novel advanced drug delivery systems: antiproliferative activity of loaded doxorubicin in human Caco-2 cells.** *Journal of Controlled Release* 2004, **94**:63– 74.
75. Hatfaludi T, Liska M, Zellinger D, Ousman JP, Szostak M, Ambrus A, Jalava K, Lubitz W: **Bacterial ghost technology for pesticide delivery.** *Journal of agricultural and food chemistry* 2004, **52**(18):5627-5634.
76. Hinz DC: **Process analytical technologies in the pharmaceutical industry: the FDA's PAT initiative.** *Anal Bioanal Chem* 2006, **384**(5):1036-1042.
77. Rathore AS, Winkle H: **Quality by design for biopharmaceuticals.** *Nature biotechnology* 2009, **27**(1):26-34.
78. Garcia T, Cook G, Nosal R: **PQLI Key Topics - Criticality, Design Space, and Control Strategy.** *J Pharm Innov* 2008, **3**:60–68.
79. Lepore J, Spavins J: **PQLI Design Space.** *J Pharm Innov* 2008, **3**:79–87.
80. Assis AJ, Filho RM: **Soft sensors development for on-line bioreactor state estimation.** *Computers and Chemical Engineering* 2000, **24**:1099-1103.
81. Veloso ACA, Rocha AI, Ferreira EC: **Monitoring of fed-batch E. coli fermentations with software sensors.** *Bioprocess and biosystems engineering* 2009, **32**:381–388.
82. Farza M, Busawon K, Hammouri H: **Simple Nonlinear Observers for On-line Estimation of Kinetic Rates in Bioreactors.** *Automatica* 1998, **34**(3):301-318.

83. Simutis R, Havlik I, Liibbert A: **Fuzzy-aided neural network for real-time state estimation and process prediction in the alcohol formation step of production-scale beer brewing.** *Journal of biotechnology* 1993, **27**:203-215.
84. Yu J: **A Bayesian inference based two-stage support vector regression framework for soft sensor development in batch bioprocesses.** *Computers and Chemical Engineering* 2012, **41**:134– 144.
85. Jenzsch M, Simutis R, Eisbrenner G, Stuckrath I, Lubbert A: **Estimation of biomass concentrations in fermentation processes for recombinant protein production.** *Bioprocess and biosystems engineering* 2006, **29**(1):19-27.
86. Wechselberger P, Sagmeister P, Herwig C: **Real-time estimation of biomass and specific growth rate in physiologically variable recombinant fed-batch processes.** *Bioprocess and biosystems engineering* 2012.
87. Sonnleitner B, Locher G, Fiechter A: **Biomass determination.** *Journal of biotechnology* 1992, **25**(1-2):5-22.
88. Kell DB, Markx GH, Davey CL, Todd RW: **Real-time monitoring of cellular biomass: methods and applications.** *Trends in analytical chemistry* 1990, **9**(6):190-194.
89. Warth B, Rajkai G, Mandenius CF: **Evaluation of software sensors for on-line estimation of culture conditions in an Escherichia coli cultivation expressing a recombinant protein.** *Journal of biotechnology* 2010, **147**(1):37-45.
90. Cannizzaro C, Gugerli R, Marison I, von Stockar U: **On-line biomass monitoring of CHO perfusion culture with scanning dielectric spectroscopy.** *Biotechnology and bioengineering* 2003, **84**(5):597-610.
91. Neves AA, Pereira DA, Vieira LM, Menezes JC: **Real time monitoring biomass concentration in Streptomyces cultivations with industrial media using a capacitance probe.** *Journal of biotechnology* 2000, **84**:45–52.
92. Davey CL, Kell DB: **The influence of electrode polarisation on dielectric spectra, with special reference to capacitive biomass measurements. I. Quantifying the effects on electrode polarisation of factors likely to occur during fermentations.** *Bioelectrochemistry and Bioenergetics* 1998, **46**:91–103.
93. Davey CL, Davey HM, Keli DB: **Introduction to the dielectric estimation of cellular biomass in real time, with special emphasis on measurements at high volume fractions.** *Analytica Chimica Acta* 1993, **279**:155-161.
94. Harris CM, Todd RW, Bungard SJC, Lovitt RW, Morris G, Keli DB: **Dielectric permittivity of microbial suspensions at radio frequencies: a novel method for the real-time estimation of microbial biomass.** *Enzyme Microb Technol* 1987, **9**(3):181-186.
95. Kiviharju K, Salonen K, Moilanen U, Eerikainen T: **Biomass measurement online: the performance of in situ measurements and software sensors.** *Journal of industrial microbiology & biotechnology* 2008, **35**(7):657-665.
96. Asami K, Yonezawa Y, Wakamatsu T, Koyanagi N: **Dielectric spectroscopy of biological cells.** *Bioelectrochemistry and Bioenergetics* 1996, **40**:141-145.
97. Jenzsch M, Simutis R, Luebbert A: **Generic model control of the specific growth rate in recombinant Escherichia coli cultivations.** *Journal of biotechnology* 2006, **122**(4):483-493.

98. Bogaerts P, Vande Wouwerb A: **Parameter identification for state estimation—application to bioprocess software sensors.** *Chemical Engineering Science* 2004, **59**:2465 – 2476.
99. Dochain D: **State and parameter estimation in chemical and biochemical processes: a tutorial.** *Journal of Process Control* 2003, **13**:801–818.
100. Wechselberger P, Sagmeister P, Engelking H, Schmidt T, Wenger J, Herwig C: **Efficient feeding profile optimization for recombinant protein production using physiological information.** *Bioprocess and biosystems engineering* 2012, **35**(9):1637-1649.
101. Lee DS, Park JM, Vanrolleghem PA: **Adaptive multiscale principal component analysis for on-line monitoring of a sequencing batch reactor.** *Journal of biotechnology* 2005, **116**(2):195-210.
102. Lennox B, Montague GA, Hiden HG, Kornfeld G, Goulding PR: **Process monitoring of an industrial fed-batch fermentation.** *Biotechnology and bioengineering* 2001, **74**(2):125-135.
103. Le H, Kabbur S, Pollastrini L, Sun Z, Mills K, Johnson K, Karypis G, Hu WS: **Multivariate analysis of cell culture bioprocess data--lactate consumption as process indicator.** *Journal of biotechnology* 2012, **162**(2-3):210-223.
104. Ferreira AP, Lopes JA, Menezes JC: **Study of the application of multiway multivariate techniques to model data from an industrial fermentation process.** *Analytica chimica acta* 2007, **595**(1-2):120-127.
105. Rhee JL, Kang T: **On-line process monitoring and chemometric modeling with 2D fluorescence spectra obtained in recombinant E. coli fermentations.** *Process Biochemistry* 2007, **42**:1124-1134.
106. Sagmeister P, Wechselberger P, Herwig C: **Information Processing: Rate-Based Investigation of Cell Physiological Changes along Design Space Development.** *PDA journal of pharmaceutical science and technology / PDA* 2012, **66**(6):526-541.
107. Ow DS, Lee RM, Nissom PM, Philp R, Oh SK, Yap MG: **Inactivating FruR global regulator in plasmid-bearing Escherichia coli alters metabolic gene expression and improves growth rate.** *Journal of biotechnology* 2007, **131**(3):261-269.
108. DeLisa MP, Li J, Rao G, Weigand WA, Bentley WE: **Monitoring GFP-operon fusion protein expression during high cell density cultivation of Escherichia coli using an on-line optical sensor.** *Biotechnology and bioengineering* 1999, **65**(1):54-64.
109. Wechselberger P, Herwig C: **Model-based analysis on the relationship of signal quality to real-time extraction of information in bioprocesses.** *Biotechnology progress* 2012, **28**(1):265-275.
110. Laemmli UK: **Cleavage of structural proteins during the assembly of the head of bacteriophage T4.** *Nature* 1970, **227**(5259):680-685.
111. O'Donnell D, Wang L, Xua J, Ridgway D, Gua T, Moo-Young M: **Enhanced heterologous protein production in Aspergillus niger through pH control of extracellular protease activity.** *Biochemical Engineering Journal* 2001, **8**:187–193.
112. Wang Y, Jing C, Yang B, Mainda G, Dong M, Xu A: **Production of a new sea anemone neurotoxin by recombinant Escherichia coli: Optimization of culture conditions using response surface methodology.** *Process Biochemistry* 2005, **40**:2721–2728.

113. Baka M, Derlinden EV, Boons K, Mertens L, Van Impe JF: **Impact of pH on the cardinal temperatures of E. coli K12: Evaluation of the gamma hypothesis.** *Food Control* 2013, **29** 328-335.
114. Song JM, An YJ, Kang MH, Lee Y-H, Cha S-S: **Cultivation at 6–10°C is an effective strategy to overcome the insolubility of recombinant proteins in Escherichia coli.** *Protein expression and purification* 2012, **82**:297–301.
115. Mejia R, Gomez-Eichelmann MC, Fernandez MS: **Membrane fluidity of Escherichia coli during heat-shock.** *Biochimica et biophysica acta* 1995, **1239**(2):195-200.
116. Glick BR: **Metabolic load and heterologous gene expression.** *Biotechnology advances* 1995, **13**(2):247-261.
117. Rao NN, Liu S, Kornberg A: **Inorganic polyphosphate in Escherichia coli: the phosphate regulon and the stringent response.** *Journal of bacteriology* 1998, **180**(8):2186-2193.
118. Hong T, Kong A, Lam J, Young L: **Periplasmic Alkaline Phosphatase Activity and Abundance in Escherichia coli B23 and C29 during Exponential and Stationary Phase.** *Journal of Experimental Microbiology and Immunology* 2007, **11**:8-13.
119. Kumar M, Upreti RK: **Impact of lead stress and adaptation in Escherichia coli.** *Ecotoxicology and environmental safety* 2000, **47**(3):246-252.
120. Andersson L, Yang S, Neubauer P, Enfors SO: **Impact of plasmid presence and induction on cellular responses in fed batch cultures of Escherichia coli.** *Journal of biotechnology* 1996, **46**(3):255-263.
121. Hewitt CJ, Nebe-Von Caron G, Nienow AW, McFarlane CM: **The use of multi-parameter flow cytometry to compare the physiological response of Escherichia coli W3110 to glucose limitation during batch, fed-batch and continuous culture cultivations.** *Journal of biotechnology* 1999, **75** 251 – 264.
122. Franchini AG, Egli T: **Global gene expression in Escherichia coli K-12 during short-term and long-term adaptation to glucose-limited continuous culture conditions.** *Microbiology* 2006, **152**(Pt 7):2111-2127.
123. Price PB, Sowers T: **Temperature dependence of metabolic rates for microbial growth, maintenance, and survival.** *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(13):4631-4636.
124. Epps DE, Wolfe ML, Groppi V: **Characterization of the steady-state and dynamic fluorescence properties of the potential-sensitive dye bis-(1,3-dibutylbarbituric acid)trimethine oxonol (Dibac4(3)) in model systems and cells.** *Chemistry and physics of lipids* 1994, **69**(2):137-150.
125. Pierucci O: **Dimensions of Escherichia coli at various growth rates: model for envelope growth.** *Journal of bacteriology* 1978, **135**(2):559-574.
126. Andersson L, Strandberg L, Enfors SO: **Cell segregation and lysis have profound effects on the growth of Escherichia coli in high cell density fed batch cultures.** *Biotechnology progress* 1996, **12**(2):190-195.
127. Sundstrom H, Wallberg F, Ledung E, Norrman B, Hewitt CJ, Enfors SO: **Segregation to non-dividing cells in recombinant Escherichia coli fed-batch fermentation processes.** *Biotechnol Lett* 2004, **26**(19):1533-1539.

128. BREMER H, DENNIS PP: **Modulation of Chemical Composition and Other Parameters of the Cell by Growth Rate.** In: *Escherichia coli and Salmonella: Cellular and Molecular Biology*. Edited by Neidhardt FC. Washington, DC; 1996: 1553–1569.
129. Batas B, Schiraldi C, Chaudhuri JB: **Inclusion body purification and protein refolding using microfiltration and size exclusion chromatography.** *Journal of biotechnology* 1999, **68**(2-3):149-158.